The University of Reading

Met Office

Numerical Weather Prediction

Results from high resolution modelling of convective events
(Stage 1 report from the Storm-Scale numerical modelling project)

Forecasting Research Technical Report No. 402
Joint Centre for Mesoscale Meteorology Report No. 140

Nigel Roberts

email: nwp_publications@metoffice.com

©Crown Copyright

Results from high resolution modelling of convective events
Stage 1 report from the Storm-Scale numerical modelling project

Date: 24th February 2003

Author: Nigel Roberts

Contents

1 Overview 4
2 Introduction 4
2.1 The aim of this report 4
2.2 The potential benefit of storm-scale numerical modelling 5
2.3 About the model setup 6
2.4 Choice of case Studies 7
2.5 Evaluation of model performance 8
3 Case study 1. 2nd to 3rd July 1999 9
3.1 Brief description of the event 9
3.2 Comparison of forecasts using different resolutions 10
3.2.1 Operational versus high-resolution forecasts 10
3.2.2 A realistic squall line 12
3.2.3 Propagation speed 13
3.2.4 Viewing the peak rainfall intensity 14
3.3 Summary of this case 15
4 Case study 2. 11th to 12th October 2000 16
4.1 Description of the event 16
4.2 Comparison of forecasts using different resolutions 16
4.2.1 Rainfall rates and precipitation structure 17
4.2.2 Rainfall accumulations 18
4.2.3 Quantitative verification against radar 20
4.3 Summary of this case 23
5 Case study 3. 3rd May 2002 25
5.1 Description of the event 25
5.2 Comparison of forecasts using different resolutions 25
5.2.1 11 UTC. Early shower development. 26
5.2.2 14 UTC. Afternoon showers 28
5.2.3 18 UTC. Evening organisation. 31
5.2.4 Mean precipitation rates over the whole period 33
5.2.5 Output of precipitation rates 35
5.3 Summary of this case 36
6 Case study 4. 29th July 2002 37
6.1 Description of the event 37
6.2 Comparison of forecasts using different resolutions 38
6.2.1 The 12 and 4 -km gridlength simulations 39
6.2.2 The 1.2-km gridlength simulations 40
6.2.3 The 4-km gridlength, 76 level run with modified convection scheme 42
6.2.4 Output of precipitation rates 43
6.3 Summary of this case 44
7 Conclusions and further work 45
7.1 Conclusions 45
7.2 Future work 45
8 References 47
9 Acknowledgements 47

1 Overview

Four case studies have been examined using high-resolution versions of the Met Office Unified Model with gridlengths of 12, 4 and either 2, 1.2 or 1 km. The case studies are all examples of convective events that produced either high rainfall rates or high rainfall accumulations (or both), yet are all very different from each other. In all four cases, the highest-resolution simulations produced better results than the other runs. The studies show that high-resolution modelling works, and has the potential to deliver improved catchment-scale forecasts of flood-producing rainfall events. Future work will consist of further case studies that examine the sensitivity to changes in model configuration and parameters and the inclusion of observations at high resolution. Model diagnostics to optimise the usefulness of model output have been presented. These will continue to be developed along with verification methods.

2 Introduction

This is the first interim report to be delivered from the Storm Scale Numerical Modelling project. The project was set-up to investigate the ability of a storm scale configuration of the Met Office Numerical Weather Prediction (NWP) model to predict flood-producing rainfall events. In this context a storm-scale model has a horizontal grid spacing of about 1km.
The reasons for attempting this now are:

(1) It is more important than ever to be able to accurately predict extreme local rainfall over small/medium sized river catchments so that adequate warning of damaging and costly floods can be issued.
(2) Current forecasts of 'flash-flood' events mostly rely on nowcasting techniques that have probably gone through the peak of their development and may have limited scope for long-term improvement. Other approaches should be examined.
(3) The Met Office has now made operational (since August 2002) a new version of its NWP model that will provide the framework for a future forecast system to be run on grids with a spacing of 1km (or shorter). Previously, this would not have been possible.

At present, NWP forecasts on a grid spacing of ~1 km could not be run in real time because of the limitations in computer resources. However, it is possible to examine the performance of such a model in a retrospective way on a set of appropriate case studies and test out possible future configurations. This is the approach that has been adopted.

2.1 The aim of this report

The primary aim of this report is to examine the results from a few initial case studies, of significant rainfall events, to determine whether high-resolution model simulations can produce sensible looking results and then see if there are any improvements over the current operational mesoscale model (grid spacing of 12 km). This will give an indication of what a 1, 2 or 4 km gridlength model is capable of in its current setup and what that implies for the future development of a storm-scale-model forecast system.

At this stage the emphasis is on making sure that the high-resolution version of the model is producing precipitation forecasts that are at least comparable with the standard 12-km forecasts and certainly not poorer. Any significant improvements to the forecasts would be very encouraging given that these particular high-resolution runs do not use any extra observations. Also, it is clear that, by increasing the resolution to the scale of individual storms, the way the model deals with clouds, convection and surface processes becomes extremely important and is the subject of ongoing research. Therefore, although the

model configurations used for these case studies were considered to be the best at the time, improvements are always being made and these results will be superseded. The report will not specifically focus on sensitivity tests involving changes in model parameters and formulation, because that will be the subject of subsequent reports – although sensitivity to parameter changes will be shown where appropriate.

The case studies were specifically chosen to be convective events that produced large rainfall totals or high rainfall rates, but were varied in character. It is envisaged that the main benefit of a storm-scale forecast system would be an improved capability to predict extreme convective events on the scale of small/medium river catchments. Assessment of forecasts was largely determined by the nature of the individual case studies. The main factors that were considered were: - rainfall accumulations, rainfall rates, convective triggering, convective decay, convective organisation, propagation speed, meteorological realism, cell sizes and spatial and temporal error. Only some of these were of interest for any given case study. Verification issues will be mentioned later.

The final part of this report will be a discussion of work that is still required within this project to develop and evaluate a storm-scale model forecast system and the necessary products for customers.

2.2 The potential benefit of storm-scale numerical modelling

The potential benefits of being able to run a forecast model routinely with a gridlength of around 1km are enormous. Many more mesoscale weather events can be properly resolved and local effects much better simulated. This should result in substantially improved forecasts and lead to new and improved forecast products.

In terms of forecasting convective storms better, one major advantage a 1km gridlength model has over the current operational 12km model is that a convection scheme is probably no longer required (or at least far less important and restricted to shallow convection).
The convection scheme is a section of code in the model that is designed to calculate changes to the gridpoint values of temperature and humidity which are due to convection that the model is unable to resolve. The problem with a convection scheme is that it is not directly part of the model dynamics formulation and is limited in the way it can represent convection. It will NOT propagate showers, produce cell life cycles, initiate new convection on outflow boundaries or develop convective organisation. Despite these problems, a convection scheme is essential in a model with a gridlength of 12km since nearly all convective clouds are smaller than the 4-5 gridlengths (48-60km) that are required to resolve them properly. If the convection scheme was not used at this resolution, forecasts might develop too much convective instability and become unstable and crash. With a 1km-gridlength model however, most deep convective clouds can be resolved. This means that features such as squall lines, inland propagation of showers, new storm-cell initiation and convective organisation (comma clouds) can be represented by the model. It is expected that this will produce much more realistic and hopefully more accurate forecasts.

Another advantage of improved resolution is that the surface terrain – mountains, hills, coastlines, forests, urban areas, lakes – can be much better represented. Storms are often thought to be tied to urban heat islands, sea-breeze convergence fronts, ascent or heating associated with hills or other local effects. It is thought that if these effects can be more accurately simulated, then many convective storms can be more accurately predicted.

Yet another benefit is that precipitation regions can be more realistically simulated with a grid spacing of ~1km. Rainfall rates on these scales will be more closely comparable with those measured by radar and rain gauges. It should be possible to make use of model output to warn of the possibility of local flash floods and the hazards associated with intense rainfall rates and then verify these forecasts.

The situation where extra resolution may not be so influential is when the errors in a forecast are fed in through the boundaries of the high-resolution model domain. This may be, for example, if a front or an upper-level vorticity anomaly is misplaced. In these circumstances, it is still possible that extra resolution will be beneficial and this needs to be examined with case studies of this nature. The impact of errors of this kind should be alleviated as new methods for introducing observations at frequent intervals from radars, profilers and satellites into high-resolution models are developed. Indeed, the instances of poorly forecast larger-scale features in the coarser-scale driving model should also be reduced as methods for incorporating observations are improved alongside improved resolution.

2.3 About the model setup

The model used is the Met Office Unified Model (Cullen et al 1997) which has a new non-hydrostatic formulation of the dynamics section. It is the non-hydrostatic nature of the model that can deal with vertical accelerations and makes it a useful tool for storm-scale modelling. The model was run with horizontal gridlengths of 12, 4, 2, 1.2 and 1km and either 38, 45 or 76 vertical levels.

Each of the case studies was firstly run on the operational mesoscale domain which has a gridlength of ~12km. These runs supplied the boundaries for the 4km-gridlength simulations that were run on case-specific smaller domains within the 12km gridlength domain. The 4km gridlength runs supplied the boundaries for the 1, 1.2 or 2km gridlength simulations that were run on case-specific smaller domains within the 4km domains. Information through the boundaries was passed from the coarser resolution to the finer resolution domains but not the other way.

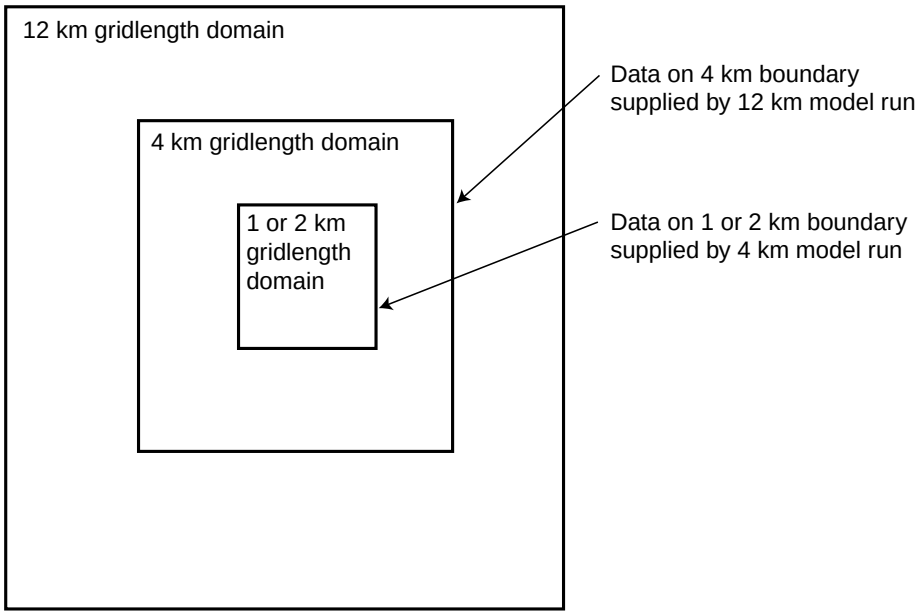


Figure 1. **Schematic diagram showing the way the domains are sequentially nested inside each other.**

The 12km gridlength model incorporated observations through the 3D-VAR data assimilation process (Lorenc et al 2000. This meant that surface observations, radiosonde soundings, aircraft, satellite, radar and profiler observations had been assimilated into the model before the start time and the model fields contained a best possible representation of the state of the atmosphere at that time (and that resolution). No further data assimilation was done on the higher-resolution domains. The 4-km gridlength starting fields simply came from an interpolation of the 12-km fields. The 1, 1.2 or 2km gridlength starting fields came from an interpolation of the 4km fields (which is in effect an interpolation from 12km). This meant that the 4, 2 and ~1-km gridlength simulations did not start out with any information at a resolution better than a gridlength of 12km. This is not ideal because it means that the high-resolution forecasts have to 'spin up' their own preferred scales of motion and it also means that it has not been possible yet to take proper advantage of high-resolution datasets like radar fields in the high-resolution domains. Ongoing research at the Joint Centre for Mesoscale Meteorology (JCMM) is addressing this difficult problem and future experiments in this project should be able to make use of that work.

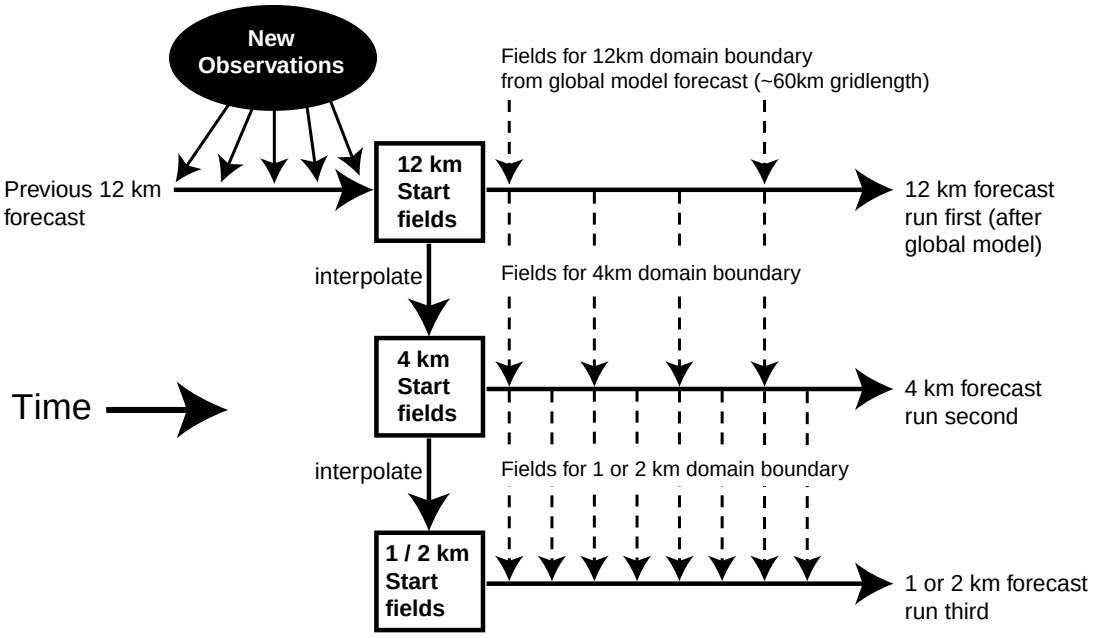


Figure 2. A schematic diagram showing the setup of the model simulations.

2.4 Choice of case Studies

Four case studies have been run. These are

1. 2-3rd July 1999. Rapidly moving severe thunderstorms and squall line with some places receiving >50mm of rain in 1 hour.
2. 11-12th October 2000. A quasi-stationary band of convective rain over Sussex and Kent that produced severe flooding.
3. 3rd May 2002. Scattered convection and thunderstorms with some organisation.
4. 29th July 2002. Isolated thunderstorms over East Anglia with flash floods.

The reason for choosing these particular case studies is because they are all different, they are all heavy rain events and also because they were associated with varying success in operational forecasts. Case 1 was well forecast, cases 2 and 3 were moderately well forecast and case 4 was poorly forecast.

The varying nature of the cases meant that different aspects of model forecast performance could be examined. The key aspects for each case were
Case 1 – Precipitation intensity, squall line organisation, propagation speed, precipitation distribution and larger-scale forcing
Case 2 – Precipitation location, rainfall accumulations and precipitation intensity
Case 3 – Convective initiation, convective decay, convective organisation, precipitation intensity and convective cell sizes
Case 4 – Convective initiation and triggering, precipitation intensity and larger-scale forcing

All of these cases were chosen because they were interesting meteorologically. There was NO attempt to pick out events that would probably lend themselves to better high-resolution simulations. In fact, this would be difficult to do, and if anything these are mostly the sort of cases where it might be expected that high-resolution forecasts would struggle to improve on the operational forecasts.

2.5 Evaluation of model performance

A meaningful evaluation of the performance of high-resolution precipitation forecasts is both difficult and extremely important. Current methods of forecast evaluation range between being completely subjective and dependent on the human eye to being completely objective and dependent on the use of quantitative verification scores. Neither is entirely satisfactory in isolation. The human eye is very good at processing images, but is easily deceived. Quantitative scores should be more reliable, but they can also be deceptive as different methods can give contradictory results.

Additional difficulties are associated with the verification of high-resolution model output. (1) Representivity errors become more important. Higher-resolution brings with it a higher spatial variability in forecasts. E.g. temperature will fluctuate more in showery conditions, precipitation rates will be less homogeneous. This will often result in poorer verification scores, even if the meteorological realism of the forecast is better. (2) No forecast model can properly resolve features that are smaller than several gridlengths. The problem with interpreting high-resolution fields is that they can look very convincing on scales that we should not believe. We therefore do not wish to evaluate the model performance of fields like precipitation on scales approaching the model gridlength. (3) Some of the errors in high-resolution forecasts do not originate in the high-resolution domain; they are a result of larger-scale errors in the flow through the boundaries. We somehow need to avoid contaminating an assessment of internal high-resolution forecast performance with these errors. However, they still count as errors when it comes to an assessment of the feasibility of a high-resolution forecast system.
Another question is what we use to verify the high-resolution forecasts against. The best way of verifying the spatial distribution of precipitation forecasts is against radar. However, radar data may not have the accuracy we would like. Rain-gauge data is more accurate but, again, representivity errors become a big problem.

The reason for mentioning these issues is to flag that they are recognised and that work is underway to investigate new methods of forecast verification that are sensible and appropriate for high-resolution modelling. As for the case studies described in this report, verification scores have been calculated when this was deemed useful. However, much of the assessment has been done by focusing on the most important aspects of each case and seeing how the different forecasts performed at reproducing those aspects. This was done by visual examination and by using measures such as storm propagation speed and average rainfall rates as appropriate. The key here was to distinguish between each case study rather than just apply an identical procedure to them all.

3 Case study 1. 2nd to 3rd July 1999

3.1 Brief description of the event

Severe thunderstorms developed over Wales and southern England in the late evening of 2nd July 1999 and tracked quickly east-northeast to clear away from the eastern coast of England by 07 UTC on the 3rd. The precipitation was very intense in places, with 44mm falling in less than 1 hour in Reading and more than 50mm falling at Heathrow Airport. Most of the rain at Reading fell in a 20 minutes period as rain rates measured from a tilting-syphon rain gauge at the University of Reading site exceeded 100mm/hr. Homes were flooded.

The radar sequence in Figure 3 shows the organisation and track of the precipitation regions. Rainfall rates averaged over the 5x5km squares displayed often exceeded 32mm/hr (though errors in quantitative radar measurements may be as high as 50% or more - A.J. Illingworth and D Harrison private communication). Between 00UTC and 03UTC the size and intensity of the rain areas did not change much as they progressed eastward, but the southern area became more organised and formed a squall line structure (figure 3(b)).

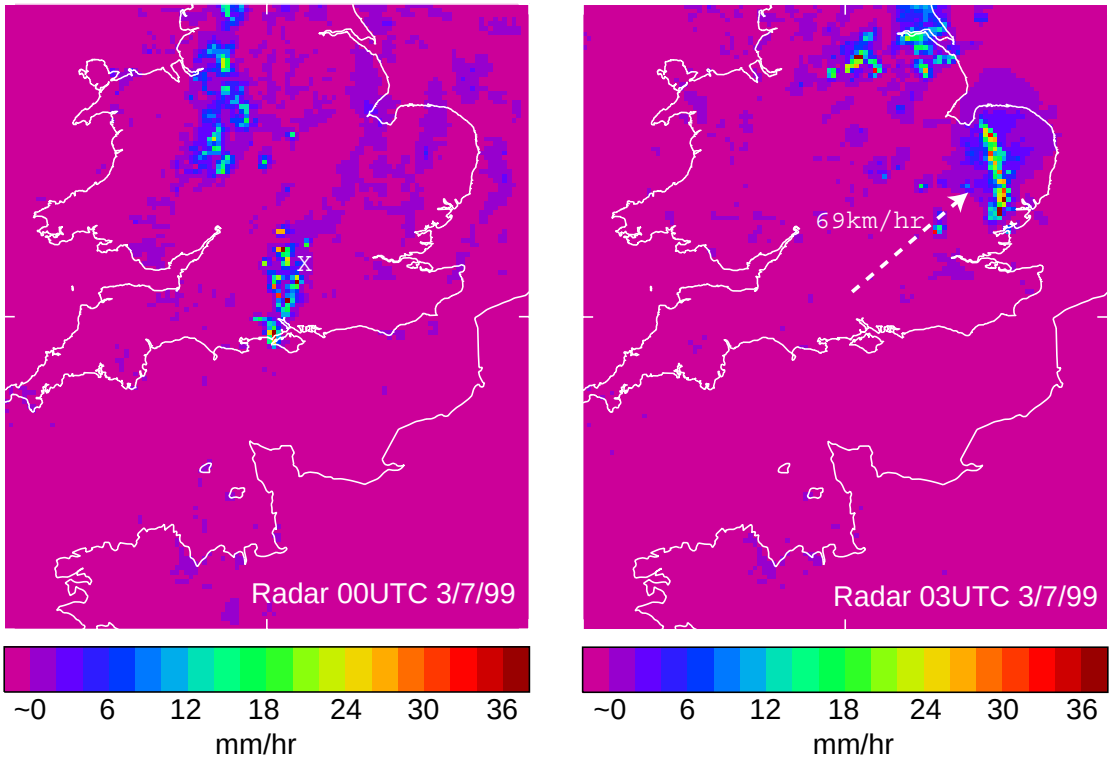


Figure 3. Rainfall rates at 00 and 03UTC 3/7/99 from the UK radar network. X marks the location of Reading. The dashed arrow shows the velocity of the southern precipitation region.

3.2 Comparison of forecasts using different resolutions

3.2.1 Operational versus high-resolution forecasts

Model simulations were run with gridlengths of 12, 4 and 2km and 45 vertical levels. The 12km forecasts (and the operational forecast) started from 09UTC 2/7/99 and ran the data assimilation process to incorporate new observations into the model fields up to 14UTC then continued with a forecast to 06UTC 3/7/99. The 4 and 2km gridlength simulations were started from 15UTC 2/7/99 and used the fields interpolated from 12km runs at that time. No extra observational information was added to the 2 and 4km simulations. The domains used are shown in Figure 4(a).

Figure 4(b) shows that the operational 12km gridlength forecast did produce precipitation in approximately the correct position with approximately the correct intensity. This was a good forecast given the propagation speed of the storms (69 km/hr). At 00UTC the model's precipitation regions were 80km too far east and 50km too far south – which equates to a timing error of only around 1 hour too soon. The fact that the operational forecast was so good in this instance means that high-resolution simulations would have to do exceptionally well to beat it.

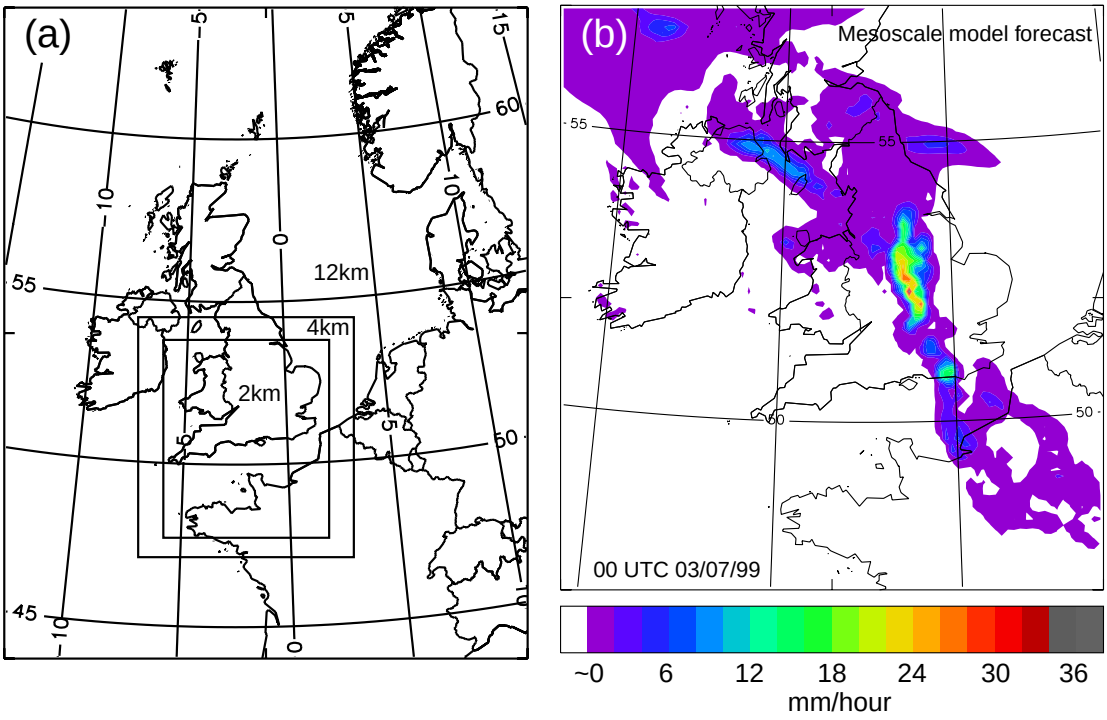


Figure 4. (a) The domains used for the 12, 4 and 2-km gridlength simulations. Rainfall rates at 00UTC 3/7/99 from a 12-hour forecast of the operational mesoscale model run on that day (shown on a sub-area of the domain).

Figure 5 shows results from 4 and 2-km simulations. There is a large difference between the runs at the same resolution that either did or did not use the convection scheme. This difference is much greater than that between the runs at different resolutions that had the same convection-scheme choice. Both the forecasts which did not use the convection scheme developed a squall line over southeast England and are more accurate forecasts in terms of the location and intensity of the precipitation than the runs with the scheme. That there should be such a marked difference between forecasts at the same resolution is significant. It is thought that, in general, a convection scheme is required for a gridlength of 4 km and not so crucial at 2km. In this particular case, the convection was so deep and

organised that the 4km run could resolve the storms without the need for a convection scheme and produced better results. This is probably unusual for the UK – in most convective situations a convection scheme is required at 4km. In light of this, work has been done to modify the convection scheme in such a way as to limit its effect so that in the intense convective events such as this one it will have less impact on the forecast than in weaker cases. A discussion of this modification will be the subject of a future report, but examples will be shown for some of the other case studies. The rest of the discussion about this case study will focus on the 4 and 2-km runs that did not use the convection scheme.

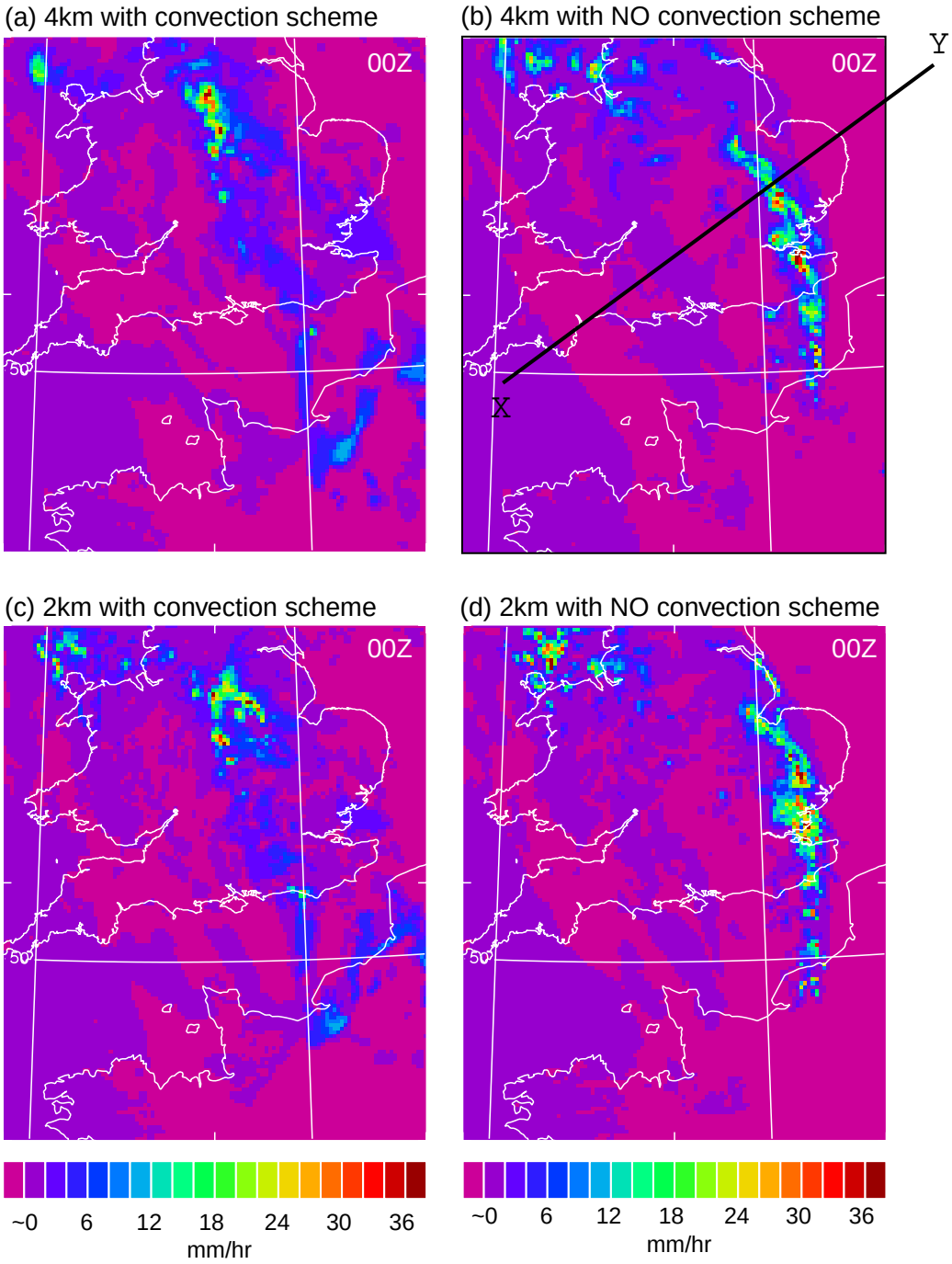


Figure 5. Rain rates at 00UTC 3/7/99 from 2 and 4km gridlength simulations that started at 15UTC 2/7/99 and followed a 6-hour 12km simulation that had data assimilation. (a) 4-km gridlength run with the convection scheme included. (b) 4 km run with no convection scheme. (c) 2-km gridlength run with the convection scheme included. (b) 2 km run with no convection scheme.

At first glance it would appear that the operational 12 km forecast is better than any of the higher resolution runs, but this not true. An appreciation of the meteorology of event is required to see why. Figure 6(a) shows schematically what was happening. The storms developed on an upper-level (700-400hPa) humidity front (curved grey line). This marked the leading edge of dry air associated with an upper-level vortex, that was overrunning a surface cold front (black dashed line). These features could be deduced from a sequence of satellite images in the Water Vapour channel along with surface observations. In all the model simulations the same features could be identified, but crucially, in all these runs the upper-level vortex was displaced further southeast and the upper-level humidity front more advanced (2-3 hours ahead) than in reality. Figures 6(b) and (c) show this difference. In the 2-km gridlength run (Figure 6(c)) the storms did develop and remain along the upper-level humidity front. However, in the 12-km forecast (Figure 6(b)) the storms were located behind the upper-level front and not along it as they should have been. The 12-km forecast appears to be better because it has compensated for the incorrect timing of the upper-level front by generating the storms in the wrong place. The 2 and 4 km simulations both developed a realistic squall line that propagated at the correct speed (section 4.2.3). They produced a more accurate representation of the meteorology of the event even though the timing error was worse.

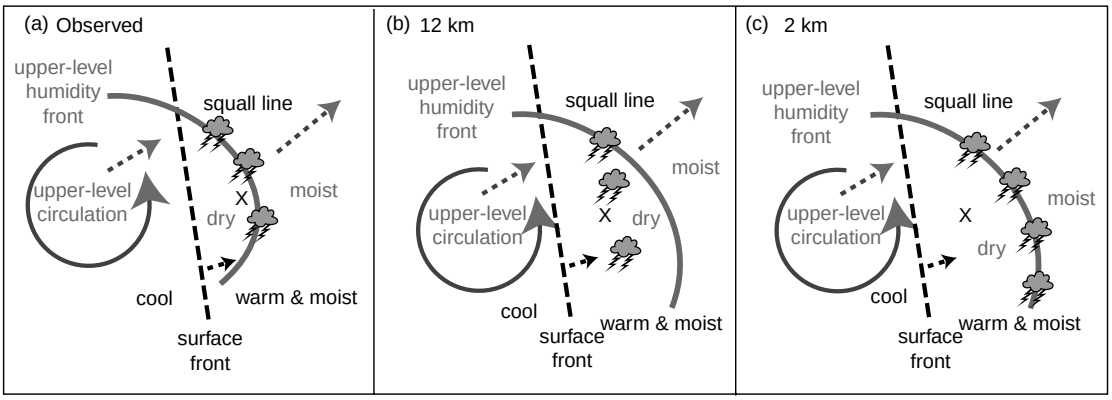


Figure 6. Schematics of a snapshot in time of the meteorological factors associated with (a) the event itself, (b) the 12 km operational forecast and (c) the 2 km simulation shown in Figure 5(d). X marks a reference location.

3.2.2 A realistic squall line

Figure 7 is included to show that the 2 and 4 km runs did produce a realistic squall line structure. It is only necessary to show a cross-section from the 4-km simulation because the equivalent from the 2-km run is very similar. The squall line occurred at the front edge of the advancing dry air at upper levels and was coherent in time over several hours. The 12 km operational forecast along with the 2 & 4 km runs which used the convection scheme did not develop this kind of structure.

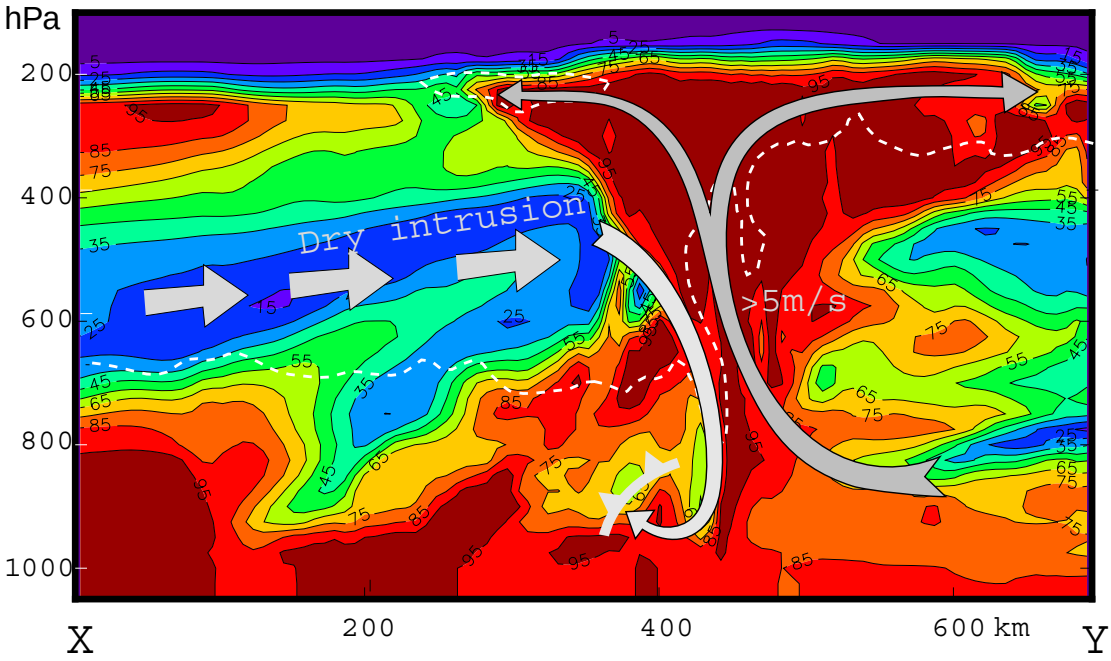


Figure 7. Cross-section along the line XY in Figure 5(b) at 00UTC 3/7/99 from the 4 km gridlength simulation shown in Figure 5(b). Colours show relative humidity (blue-green = dry, brown-orange = cloudy/moist). The white dashed line shows zero flow along the section. The fat white arrows indicates the flow of dry air behind the squall line. The grey arrow shows the updraught into the squall line, the white arrow shows the downdraught.

3.2.3 Propagation speed

Rainfall rates at 00UTC from the runs shown in Figure 5(b) and (d) have been averaged to a 12 km grid and compared with a 12 km gridlength forecast that was run without the convection scheme (Figure 8). The averaging was done to give a like for like comparison. In general, to run at 12 km without a convection scheme is not a good idea, but because of the nature of this event it was tried and even gave a reasonable forecast.

A big difference between the three simulations at 00UTC was the location of the line of storms. The higher the resolution the further east the band of precipitation occurred (Figure 8(d)). The rainband positions at 00UTC were different because the propagation speeds through the forecasts were different (Figure 8(e)). The 4 km and 2 km runs propagated the squall line at roughly the same speed as the observed event, but the 12km run was much slower. A gridlength of at least 4 km was necessary to generate a squall line that travelled at close to the correct speed. Again, this indicates that the higher resolution was required to develop a squall line that behaved in a realistic way. The precipitation band also tended to develop further north with increasing resolution, which suggests that resolution played a part in modifying the larger-scale pattern as well as local structure.
This case has not been run with a shorter gridlength than 2 km so it is not known whether the propagation speed would continue to increase with a gridlength of say 1 km, or indeed, if the precipitation would move further north. Further experiments are required to answer these questions.

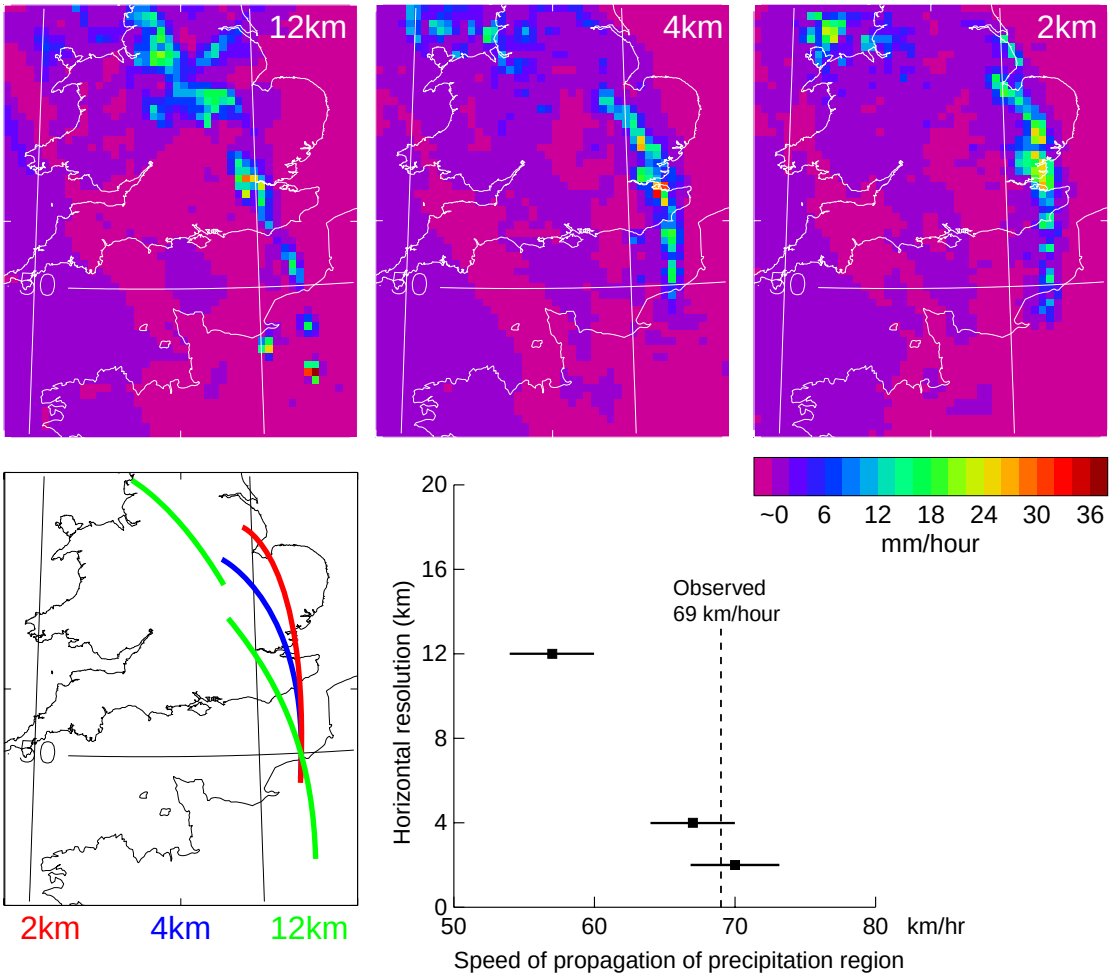


Figure 8. (a), (b) and (c) The precipitation rates at 00UTC 3/7/99 from 12, 4 and 2km resolution simulations starting at 15UTC 2/7/99 with the convection scheme switched off. (following a 6-hour 12-km forecast with data assimilation). The 4 and 2km rain rates have been averaged to 12km resolution to give a like for like comparison. (d) The lines show the axis of the 12, 4 and 2km rainbands from (a), (b) and (c). (e) A graph of propagation speed against horizontal gridlength.

3.2.4 Viewing the peak rainfall intensity

High-resolution simulations produce extra detail that should be exploited. Figure 9 shows a way of presenting peak rainfall rates in a way that makes use of the high-resolution output to give an indication of the regions at risk of high rates but removes the noise at the scale of a gridlength. Rain rates exceeded 80 mm/hour in both the 2 and 4km simulations. Such intensities are very plausible for this event. The tilting syphon rain gauge at the Reading University site recorded rain rates in excess of 150 mm/hour and the 55mm in 1 hour at Heathrow indicates that rain rate must have been extremely high there too. However, rates over 80 mm/hour in the 4-km run may well be too high since it is an average over 4x4km squares. The equivalent picture generated from the radar network (5x5-km gridlength) (Figure 9(c)) shows rates exceeding only 30 mm/hour. These lower rates are probably partly due to the averaging of the radar data to a 5km gridlength before extracting peak values, but may also indicate an underestimate in the radar measurements. The 2-km rain rates are more realistic because they represent smaller squares and if averaged up to a 5-km grid are closer to the radar. Serious evaluation of the accuracy of

rainfall intensities from high-resolution forecasts really requires a comparison with high-resolution radar data (~ 1 or 2 km squares) which is becoming available (but was not for this event) in conjunction with rain-gauge measurements.

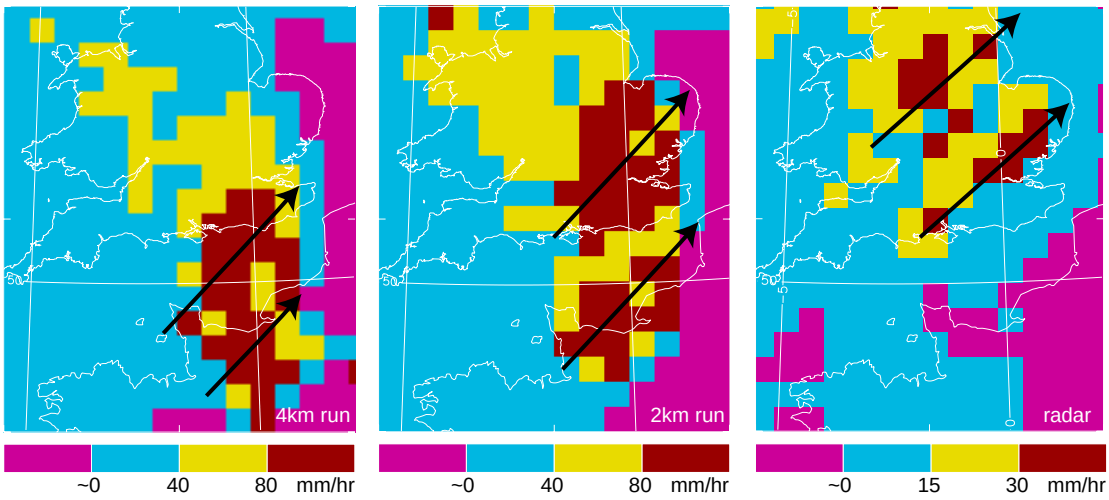


Figure 9. (a) 4km and (b) 2km resolution runs with the convection scheme switched off starting at 15UTC. Maximum precipitation rates to occur in 40x40km squares in the period 21UTC 2/7/99 to 00UTC 3/7/99 sampling at 10-minute intervals. (c) Radar, maximum precipitation rates to occur in 40x40km squares in the period 23 UTC 2/7/99 to 03 UTC 3/7/99. Arrows show the tracks of high rain-rate regions.

Figure 9 also reveals that the 2-km simulation tracked the intense precipitating areas over 100km further north than in the 4-km forecast. The 2-km forecast compares much better with the precipitation tracks shown in the radar picture. This way of presenting the output has revealed a significant improvement in the 2 km run over the 4 and 12 km runs which was not immediately evident from the snapshots shown in Figures 5 and 8.

3.3 Summary of this case

The case was of rapidly moving intense thunderstorms that developed into a squall line. The model was run with gridlengths of 12, 4 and 2 km. The operational 12 km forecast of this event was good, but this was partly due to the forecast generating the storms in the wrong place relative to the dynamical forcing in the model and by so doing compensating for a timing and spatial error in the model fields. The 4 and 2-km forecasts developed a realistic looking squall line when the convection scheme was switched off. In terms of the criteria that have been examined to assess forecast performance in this case, the 2km forecast with no convection scheme was consistently the best.

Criteria for forecast assessment	The best forecast
Meteorological realism	2 or 4 km gridlength, no convection scheme
Propagation speed	2 km gridlength, no convection scheme
Rainfall rates	2 km gridlength, no convection scheme
Consistency with dynamical fields	2 km gridlength, no convection scheme
Precipitation track	2 km gridlength, no convection scheme
Timing	12 km operational forecast – but from compensating errors

4 Case study 2. 11th to 12th October 2000

4.1 Description of the event

A quasi-stationary band of rain remained over East Sussex and Kent for over 18 hours. Embedded convective cells propagated along this band throughout the period. Extremely high rainfall accumulations were measured with several sites recording accumulations in excess of 100mm. The highest known rainfall accumulation was at Plumpton in East Sussex, where 143.8mm fell within 24 hours. Severe flooding occurred, especially within the catchments of the River Uck at Uckfield and the River Ouse at Lewes. This was probably the most dramatic event in the record-breaking wet Autumn of 2000.

Figure 10(a) shows an instantaneous radar picture of the rainband over southeast England in which localised rainfall rates exceeded 16 mm/hour (over 5x5km areas). This picture is typical of the situation over many hours. Accumulations from the radar exceeded 70mm in places in the 15-hour period shown in Figure 10(b). The most critical aspect of this event was the very high rainfall accumulations that occurred within an area only 40 km across. The main aim of this case study was to examine the ability of the model simulations to generate comparable accumulations in the correct place

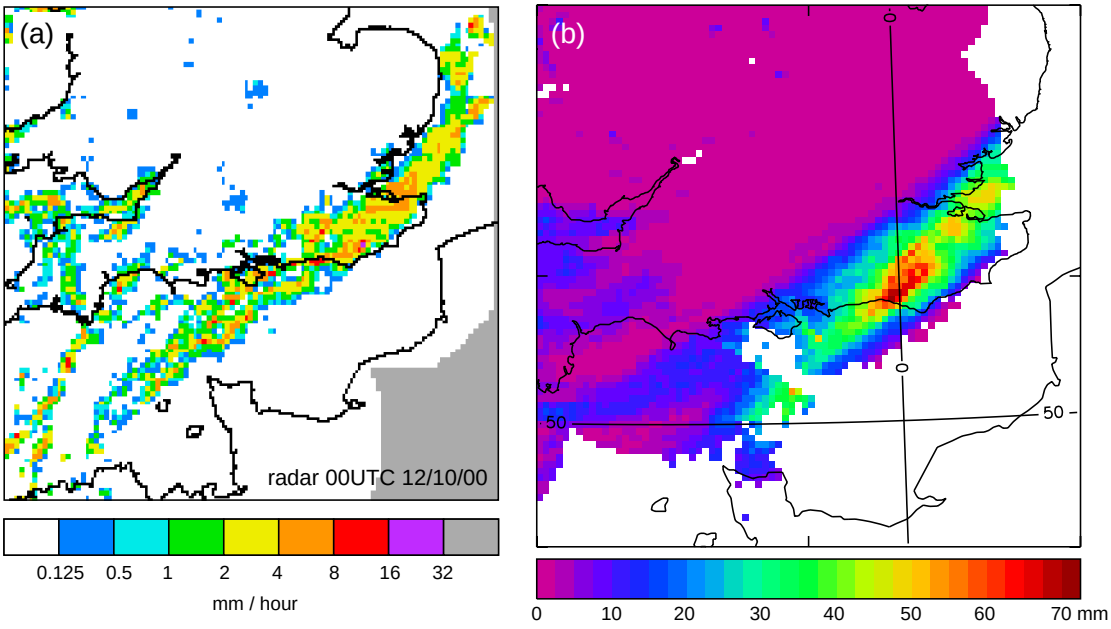


Figure 10. UK radar network with NIMROD quality control. Pixels are 5x5 km. (a) Rainfall rates at 00 UTC 12/10/00. (b) Rainfall accumulations between 19 UTC 11/10/00 and 10 UTC 12/10/00.

4.2 Comparison of forecasts using different resolutions

Model simulations were run with gridlengths of 12, 4 and 2km. The forecasts started from 18 UTC 11/10/00 and ran for 18 hours ending at 12 UTC 12/10/00. The 4 and 2km gridlength simulations used the fields interpolated from 12km at 18 UTC 11/10/00. No extra observational information was added to any of the simulations.

4.2.1 Rainfall rates and precipitation structure

Figure 11 shows output from 12, 4 and 2-km gridlength forecasts. All of these forecasts ran with the convection scheme switched on. Other experiments have been performed with the convection scheme switched off or modified, but are not presented here because the trends that emerged were essentially the same, even though individual forecasts at the same resolution did differ.

As would be expected, the 2-km forecast developed the finest structure in the precipitation pattern and the highest rainfall rates (> 16 mm/hour). Most striking though, is how similar all of these pictures are in terms of the overall distribution of rainfall. All three have generated a realistic looking band of precipitation, although the main precipitating cells are 20-50km too far to the northwest. It might be assumed from this single snapshot that the extra resolution has done little to improve the overall forecast.

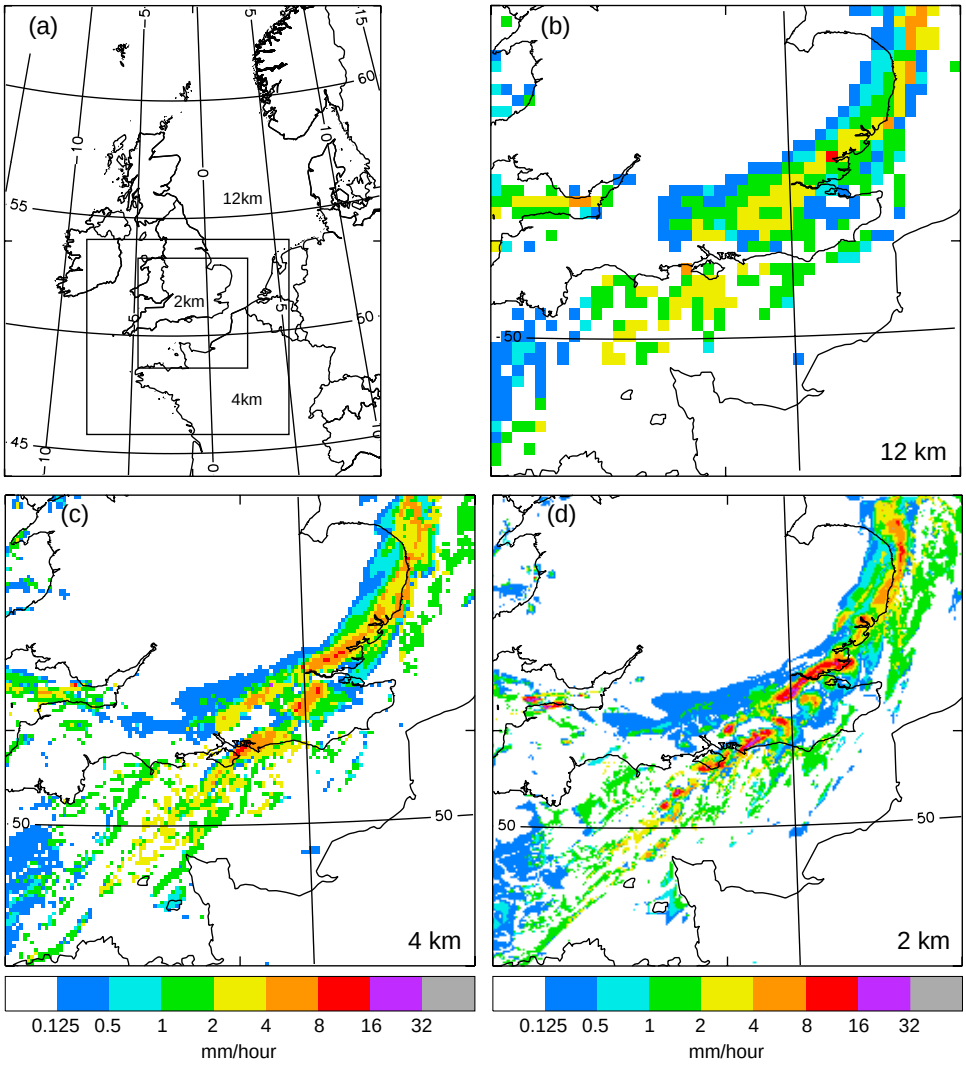


Figure 11. (a) The domains used for the 12, 4 and 2km resolution simulations. (b), (c) and (d) Rainfall rates at 00 UTC 12/10/00 from 12 km, 4 km and 2 km gridlength forecasts that started at 18 UTC 11/10/00 (shown on the 2km gridlength domain).

For a more like for like comparison, it is worth examining the rainfall rates from the 2 and 4-km forecasts when averaged to the resolution of the radar. This is shown in Figure12. The 2-km forecast still has rates greater than 16 mm/hour when averaged to 5 km, but the 4-km forecast has only a few pixels greater than 8 mm/hour. Comparison with the radar picture at 00 UTC (Figure 10(a)) suggests that the 2km is slightly too intense and the 4 km

somewhat too weak. Examination of several other times showed the 2-km to be the closest to the radar. (Note – care should be taken when averaging rainfall rates because the results can be dependent on the rainfall distribution). Compared to the radar, both the 2 and 4 km runs produce precipitation cells that were slightly too large, although, once again, the 2km forecast was the more realistic.

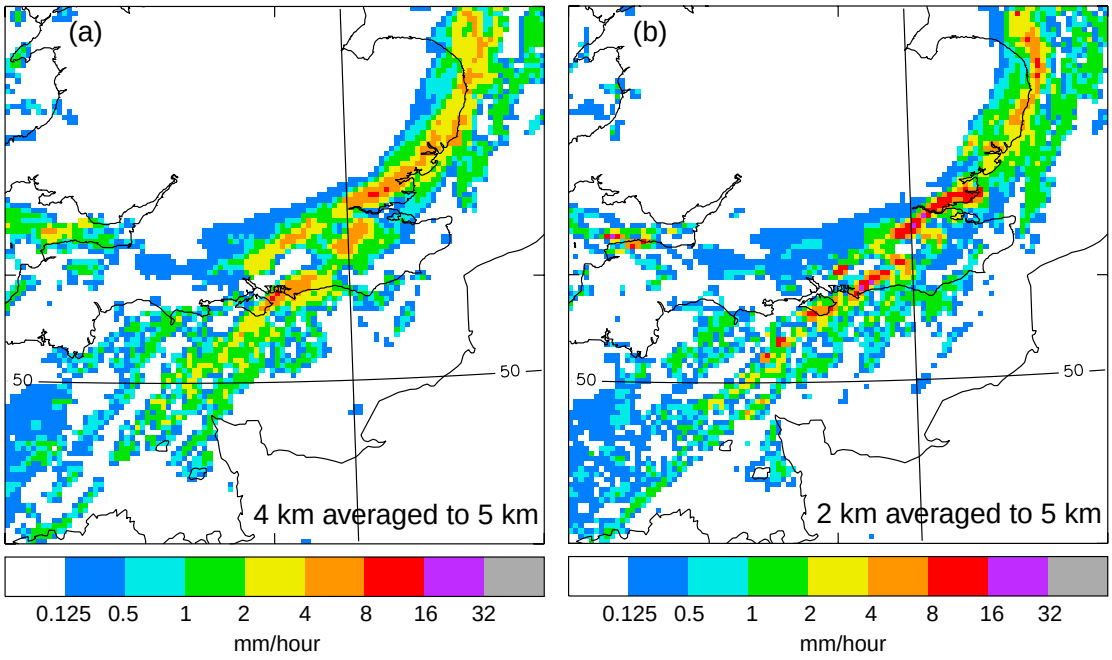


Figure 12. Rainfall rates at 00 UTC 12/10/00 averaged to a 5x5-km grid. From (a) 4 km and (b) 2 km gridlength forecasts that started at 18 UTC 11/10/00.

4.2.2 Rainfall accumulations

Accumulations from the 12, 4 and 2 km forecasts are shown in Figure 13. The 2 and 4 km output has been averaged to 12 km to give a fair comparison between all three resolutions. The 12 and 4-km runs are very similar; they both produced a band of accumulations over 30 mm through southeast England that was located 40 km to the northwest of the observed location. Both also have some higher accumulations along the south coast due to extra triggering of the convection scheme along the coast. This is unrealistic, though less of a problem at 4 km and even less at 2 km.

The major improvement comes from the 2 km run. The accumulations are much higher (>50 mm) and therefore closer to the radar accumulations (>70 mm). This may not be very surprising, since a 2 km forecast should produce higher rain rates and in turn should produce higher accumulations. Indeed, if the only benefit a higher resolution forecast gave was to increase rain rates and accumulations, it should be possible to use statistics to modify coarse resolution output accordingly and negate the need for extra resolution. However, the real improvement that comes from this 2-km forecast is the shift of the band of high rainfall totals some 20 km to the southeast. Half of the displacement error has been removed. This is remarkable, if we consider that all the runs started from the same model fields, no extra observations were added and the rainband itself is not a local phenomena. It would appear that in resolving the smaller-scale rainfall cells better, the 2km simulation is able to improve the large-scale pattern.

The peak rainfall accumulations from the 2-km run are also presented in Figure 13(d) as the maxima to occur in 12x12 km squares. The reason for doing this is to present the high-resolution information from the forecast, but not on the grid-scale of the model as spatial

accuracy is poor over a few gridlengths or less. The peak totals exceeded 70mm – just like the radar. This picture would have given a very good indication of the threat of extremely high rainfall amounts in the region where flooding occurred.

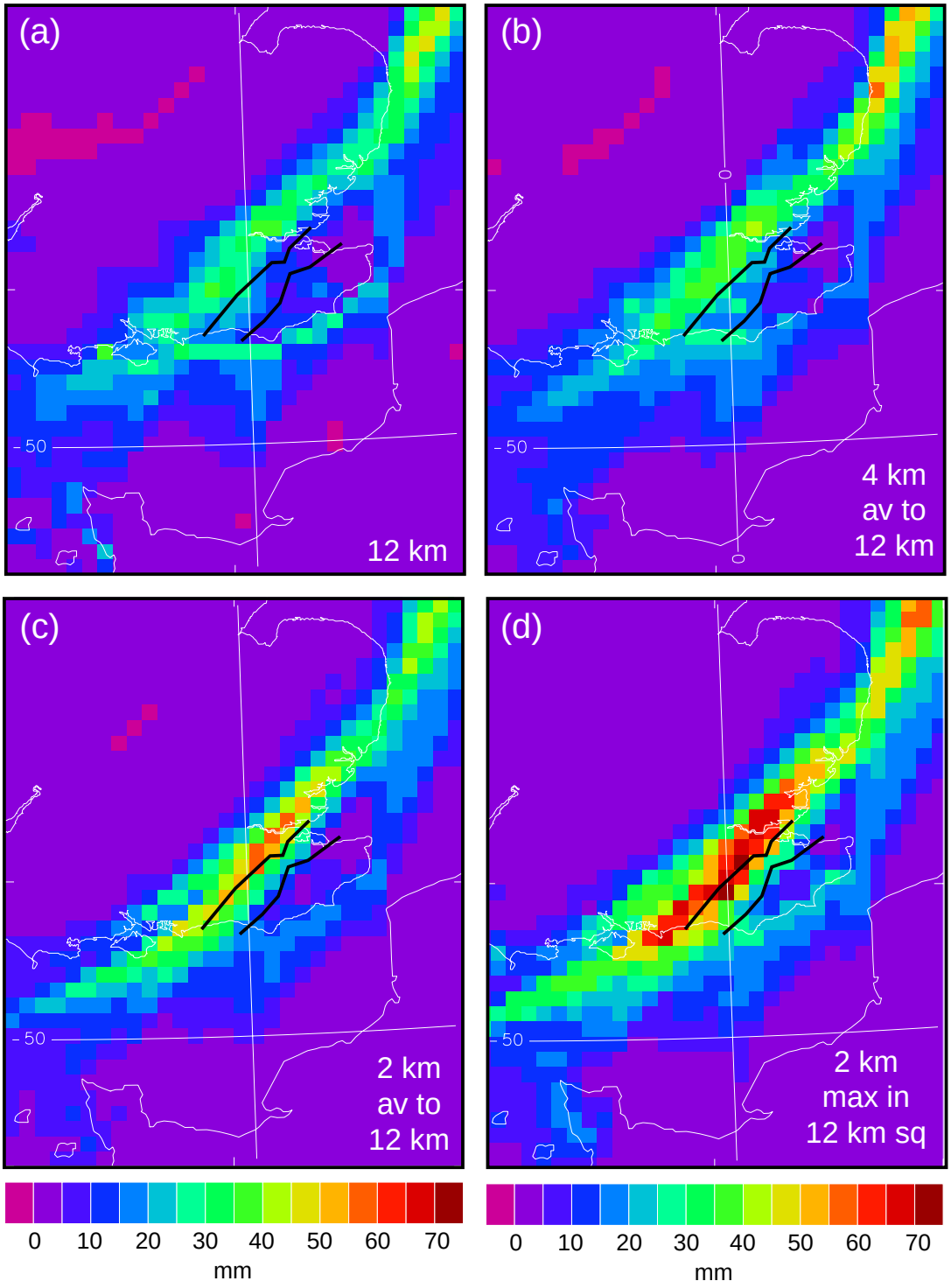
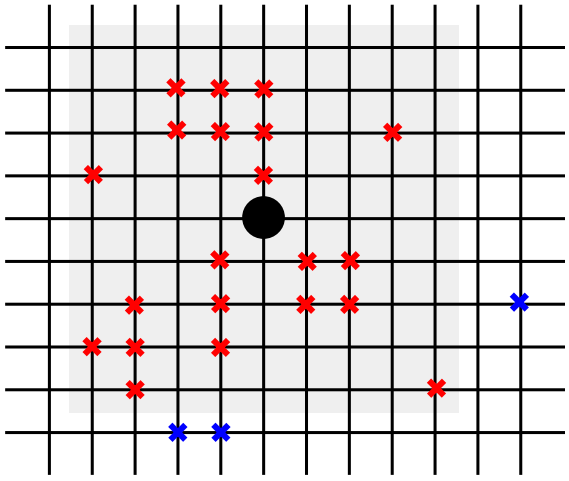


Figure 13. Rainfall accumulations over the period 18 UTC 11/10/00 to 12 UTC 12/10/00. (a) 12-km gridlength forecast. (b) & (c) 4 and 2 km forecast averaged to the 12 km grid. The black lines enclose the area where accumulations from radar exceeded 45 mm.

4.2.3 Quantitative verification against radar

Visual examination gives the impression that the 2-km gridlength run gave the best forecast, but it would be much more satisfactory if quantitative evaluation could be used to back this up. To do this we need to adopt an approach that attempts to deal with some of the problems associated with evaluating precipitation forecasts. Firstly, it is not wise to do a gridpoint against gridpoint verification of high-resolution model output because we know that detail at the grid-scale is not believable. We do not expect a 1-km forecast to get the locations of showers accurate to 1 km. Secondly, it is important to be able to determine the length scales over which a forecast is accurate. If a forecast misplaces a flood event by 20km over a river catchment of say 60x80km then the forecast is useful – if the catchment is much smaller the forecast may not be so valuable. Thirdly, if we can verify a product that customers use rather then raw output, then the purpose of evaluation becomes much more worthwhile and the requirements of customers can be fed into the verification process. Fourthly, the model output and validating data need to be processed in an identical way. Fifthly, appropriate verification scores should be used.
A simple method will be presented here. It is a first stab at tackling the verification problems that high-resolution modelling introduces. Further work is necessary and underway.

The model accumulations are verified against radar accumulations because of the spatial coverage of the radar. At this early stage, errors in the radar are not taken into account and rain gauge data is not used. The first step was to average or interpolate the model output from different resolution forecasts along with radar data to a 5-km gridlength domain. Model and radar accumulations were then converted to 'probabilities' of specified accumulations occurring. This is simply done by finding the number of points within a square surrounding each point that have an accumulation over a certain threshold.



✖ Point at which the precipitation accumulation is above a threshold in the defined square

✖ Point at which the precipitation accumulation is above a threshold outside the defined square

For gridpoint ●

The number of ✖ within defined 9x9 (81) gridpoint square surrounding that gridpoint = 21

The proportion of gridpoints with a precipitation accumulation over the defined threshold within the defined area (and within a defined time period) is

21/81 * 100 = 25.9%

This is repeated for every gridpoint in the domain

Over a period of 15 hours the precipitation accumulation at each gridpoint was found. For each gridpoint, a square was centred over that gridpoint and the proportion of other gridpoints within that square that had a precipitation accumulation over a specified threshold was found and output as a percentage. Several different sized squares and accumulation thresholds were used.

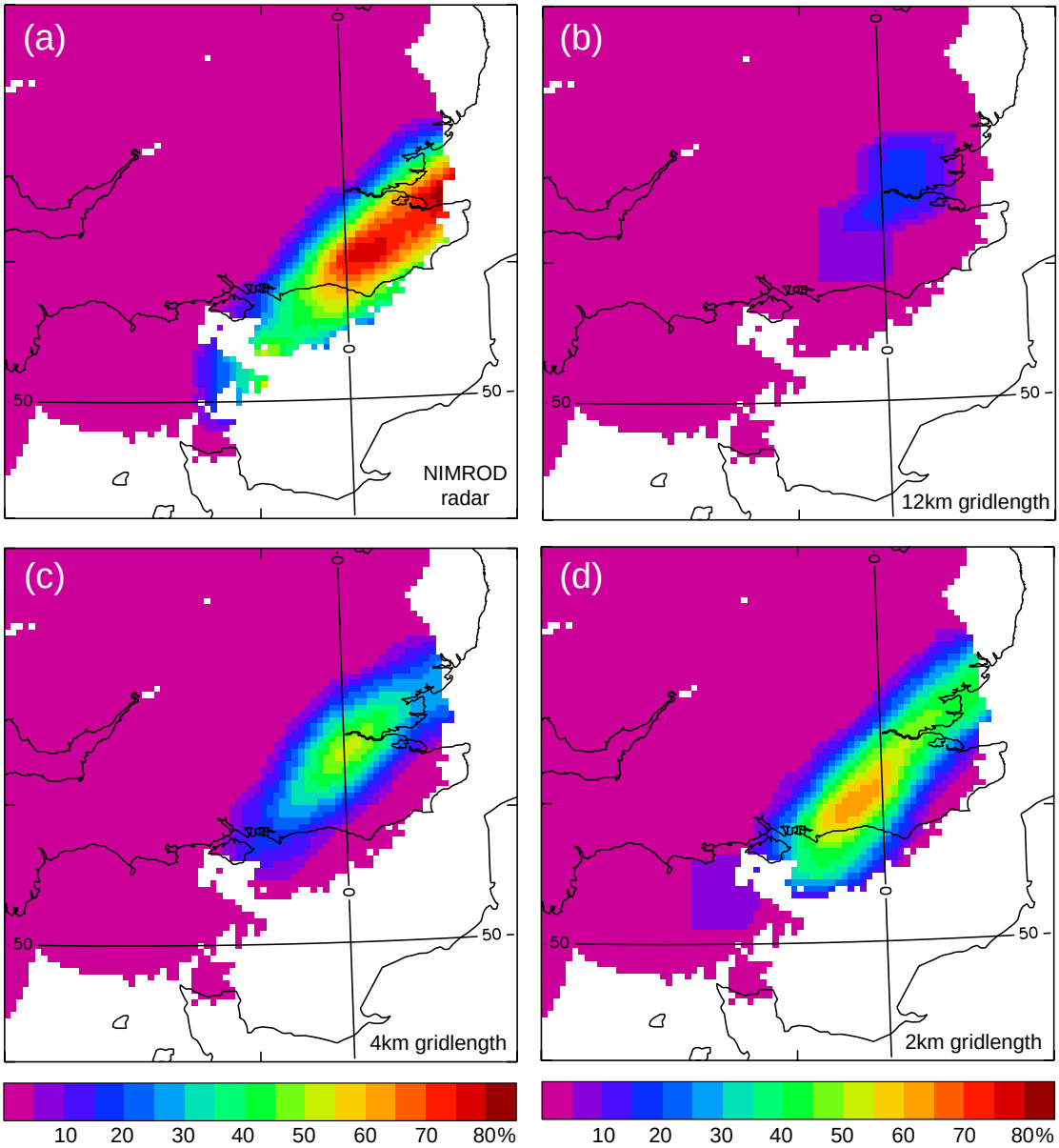


Figure 14. 'Probabilities' of rainfall accumulations exceeding 30 mm within a 15 hour period from 19 UTC 11/10/00 to 10 UTC 12/10/00 using squares of width 30km from (a) NIMROD quality controlled radar data, (b) 12 km gridlength forecast from 18 UTC 11/10/00, (c) 4 km gridlength forecast from 18 UTC 11/10/00 and (d) 2 km gridlength forecast from 18 UTC 11/10/00. A description of how this figure was generated is given in the text.

Figure 14 shows the result of using a rainfall accumulation threshold of 30 mm and squares of width 30km. The pictures reinforce the view that the 2-km forecast is closest to the radar in both magnitude and position and that the 12-km is the poorest. This kind of approach to generating information from forecasts may be a useful way of presenting high-resolution model output, since the magnitudes of the accumulations are retained

while the spatial information is selectively smoothed according to the square size. Adjustment of the square sizes or thresholds could be used to tailor the product to customer requirements, catchment-area sizes or the inherent accuracy of the model. The accuracy of the model could be determined by calculating verification scores for different sized squares and finding the smallest square that gives acceptable results over a large number of events. The scores should improve as squares become larger than displacement errors in the forecasts.

The verification required thresholds to be used for the probabilities so that categorical 'yes' or 'no' occurrences could be defined. ' Yes' for a probability above the threshold and 'no' for a probability below. Then contingency tables were set up such that: -

Hit =	forecast probability > threshold, and radar probability > threshold
Miss =	forecast probability < threshold, and radar probability > threshold
False alarm =	forecast probability > threshold, and radar probability < threshold
Correct negative =	forecast probability < threshold, and radar probability < threshold

Then the odds ratio (see below) was calculated for each required combination of accumulation and probability thresholds.

Forecast probability > threshold	Observed probability > threshold: Yes	Observed probability > threshold: No
Yes	hits	misses
No	false alarms	correct negatives

$$\text{Odds Ratio} = \frac{\text{hits x correct negatives}}{\text{misses x false alarms}}$$

The odds ratio measures the ratio of the odds of making a hit to the odds of making a false alarm. A description of the odds ratio and its value in the verification of meteorological forecasts is given by Stephenson 2000. Often the logarithm of the odds ratio is used instead of the original value and was in this case. The Log odds ratio has a range from minus infinity to infinity, 0 indicates no skill. A score of 1 indicates 10:1 odds of getting a forecast right; a score of 2 indicates 100:1 odds of getting a forecast right.

Results

The main result is that the odds ratio scores for the 2-km gridlength forecast were consistently better than for the 12 and 4-km forecasts for all square sizes. The 4 km forecast is generally better than the 12 km. Figure 15 shows that the 2 km forecast is by far the best for both an accumulation threshold of 30 mm in conjunction with a 10% probability threshold and an accumulation threshold of 20 mm in conjunction with a 40% probability threshold. Other aspects of the graph are also worth noting.

The higher accumulation threshold of 30 mm gave better scores than the 20-mm threshold when the squares were large but worse scores when the squares were small. This is largely because the 30 mm accumulations were less common than 20 mm and it was

therefore less likely for 30 mm forecast and radar pixels to occur in the same square when the squares were small, but more likely for them to be entirely contained in the same square as the squares became larger. A dip occurs in the 20mm/40% threshold (continuous) lines at all resolutions. This is related to the displacement error in the forecast and is wider for a larger displacement error. For the 12 and 4-km gridlength forecasts the dip means that the 5-km 'radius' scores are not beaten until the 'radius' is increased to around 30-35 km. For the 2-km gridlength forecast the dip is narrower and the 5-km 'radius' score is beaten at a 'radius' of 20 km. These results agree with the visual impression that the 2km forecast had the smallest displacement error and the better scores are not just due to higher accumulations being produced.

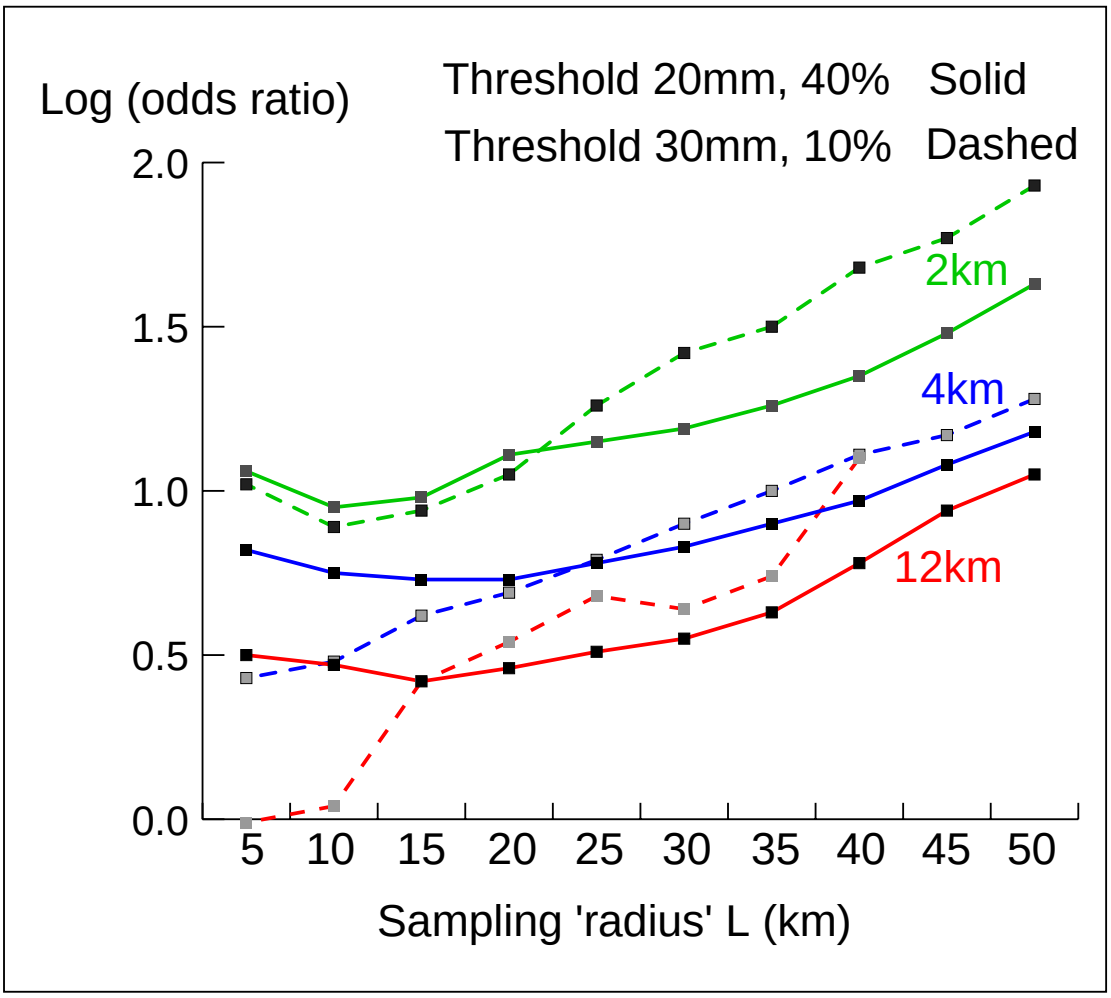


Figure 15. A graph of the Log odds ratio against square width L for 12, 4 and 2 km gridlength forecasts of rainfall accumulations compared with radar data. A description of how this graph was obtained is given in the text.

4.3 Summary of this case

The case was of a quasi-stationary convective rainband that produced extremely high rainfall accumulations within a 24-hour period over parts of East Sussex and Kent. The model was run with gridlengths of 12, 4 and 2 km; all the runs shown included the convection scheme. All of the forecasts produced a band of precipitation throughout the period. The 2-km forecast was the most accurate in all aspects examined. It produced the most realistic magnitudes of rain rates and rainfall accumulations as well as producing a

significantly better forecast of the location of the high accumulations. An objective evaluation of the forecasts using the odds ratio also showed the 2km forecast to be clearly the best.

Criteria for forecast assessment	The best forecast
Convective cell sizes	2 km gridlength
Magnitude of rainfall accumulations	2 km gridlength
Rainfall rates	2 km gridlength
Location of high rainfall accumulations	2 km gridlength

5 Case study 3. 3rd May 2002

5.1 Description of the event

Scattered showers triggered in the late morning and then developed widely over England and Wales during the day. They propagated southwards in a gentle northwesterly flow and gradually became more organised, forming an organised band over southern England in the evening before dying out overnight (Figure 16). Many of the showers were small but locally very intense with lightning. This case is very different to the others because of the scattered nature of the convection and its organisation into larger clusters. It was chosen for those reasons and also because of the inherent difficulty in simulating such small heavy showers even if using a gridlength of 1 km. The main concerns were how well the forecasts could reproduce the timing of the shower formation, the intensity of the precipitation and the organisation.

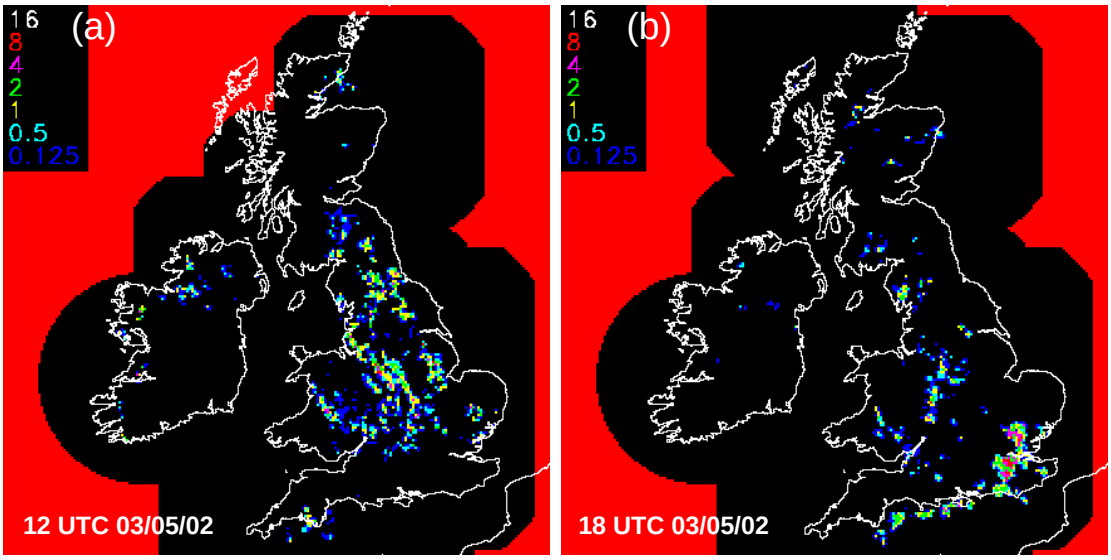


Figure 16. Radar images for 12 and 18 UTC 03/05/02.

5.2 Comparison of forecasts using different resolutions

Model simulations were run with gridlengths of 12, 4 and 1km and 38 or 76 vertical levels. The 12 and 4 km forecasts started from 01 UTC 03/05/02 and ran for 17 hours ending at 18 UTC. The 4-km gridlength simulations used the fields interpolated from 12km at 01 UTC. The 1-km gridlength runs started at 09 UTC and ran for 9 hours ending at 18 UTC. They used fields interpolated from 4-km forecasts at 09 UTC. No extra observational information was added to any of the simulations. The domains used are shown in Figure 17.

Five different model runs are examined here:

1. 12 km gridlength, 38 vertical levels, convection scheme switched on
2. 4 km gridlength, 76 levels, convection scheme switched on
3. 4 km gridlength, 76 levels, convection scheme switched off
4. 4 km gridlength, 76 levels, convection scheme modified to limit its impact
5. 1 km gridlength, 76 levels, convection scheme switched off

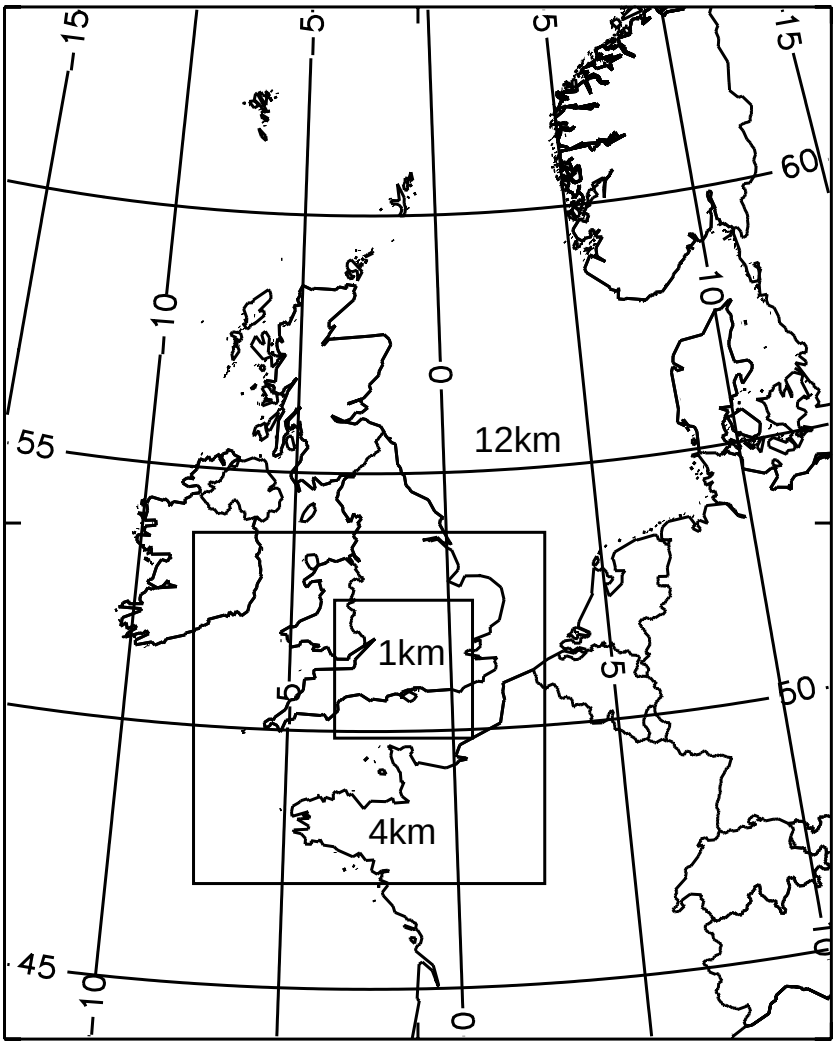


Figure 17. The domains used for the 12, 4 and 1-km gridlength simulations

Model output is presented for three times, 10, 14 and 18 UTC 03/05/02, to investigate the ability of the forecasts to reproduce the transition from the initial development through to an organised band.

5.2.1 11 UTC. Early shower development.

Rainfall rates at 11 UTC from the five forecasts and the radar are shown in Figure 18. At this time showers had developed in the north of the 1-km gridlength domain. They were small and a few had become heavy but when averaged to 5km they all appeared to be light. Some showers were forming in situ, whilst others had advected into the region from the north. Each of the forecasts had different characteristics that will now be described.

12 km with convection scheme – A few weak showers had developed, but not as widely or as far southeast as they should have. All of the precipitation came from the convection scheme.

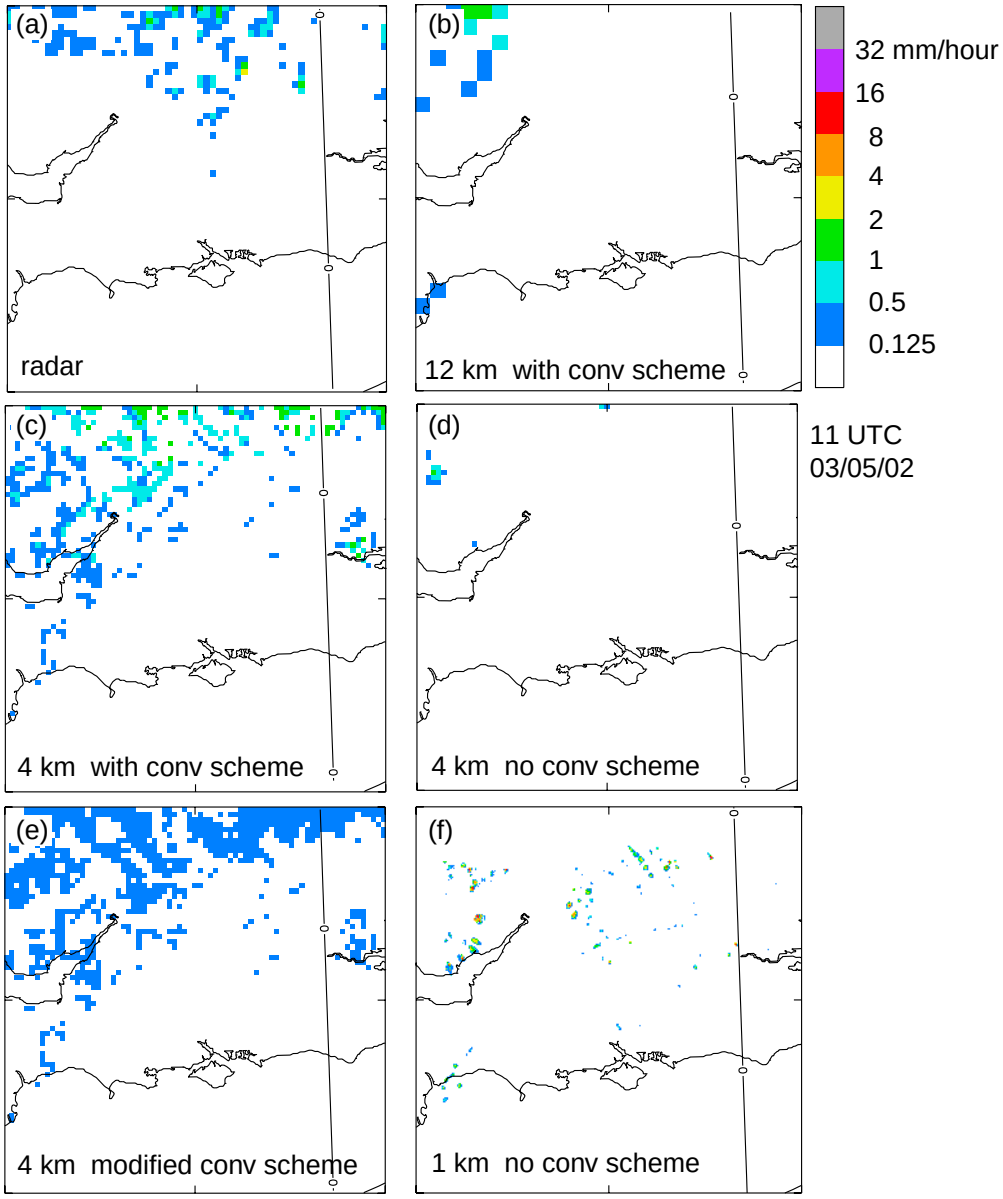


Figure 18. (a) Radar picture at 11 UTC 03/05/02. (b) to (f) Rainfall rates at 11 UTC from five different forecasts as indicated in the figure and in the text.

4 km with convection scheme – Showers had developed widely and were much more extensive than in the radar picture. The intensities are reasonable. All the precipitation came from the convection scheme.

4 km with no convection scheme – Hardly any showers had formed. The resolution was not high enough to allow triggering at that time without a convection scheme to do it.

4 km with a modified convection scheme – The rainfall pattern very closely matched the 4 km forecast with the convection scheme but the rainfall rates were uniformly less than 0.5 mm/hour. All the precipitation came from the convection scheme. This shows the effect of limiting the impact of the convection scheme at this stage.

1 km with no convection scheme – A speckling of small showers had developed. Some of the showers were already heavy (> 8mm/hour). It is difficult to compare this with the radar because of the difference in resolution of the two pictures so the model output has been

averaged to the same 5km grid as the radar and shown in Figure 19. Now it is apparent that the size and intensities of the showers were roughly comparable, though the model had some locally more intense showers even after averaging. There was however, a major difference in the location of the shower development. The 1-km forecast had no showers in the northernmost 30 km of the domain north of the dashed line). The main reason for this is that the showers have not had time to develop (or "spin up") in the flow that has recently advected through the northern boundary of the domain. This is an important issue in high-resolution modelling. The characteristics of the 4-km forecast that supplied the boundary conditions are crucial here. An investigation of this is ongoing and will be discussed in a later report. It is also worth pointing out that the 4km simulations with no convection scheme suffer in the same way, it is just that the edges of the 4-km domain have not been shown here.

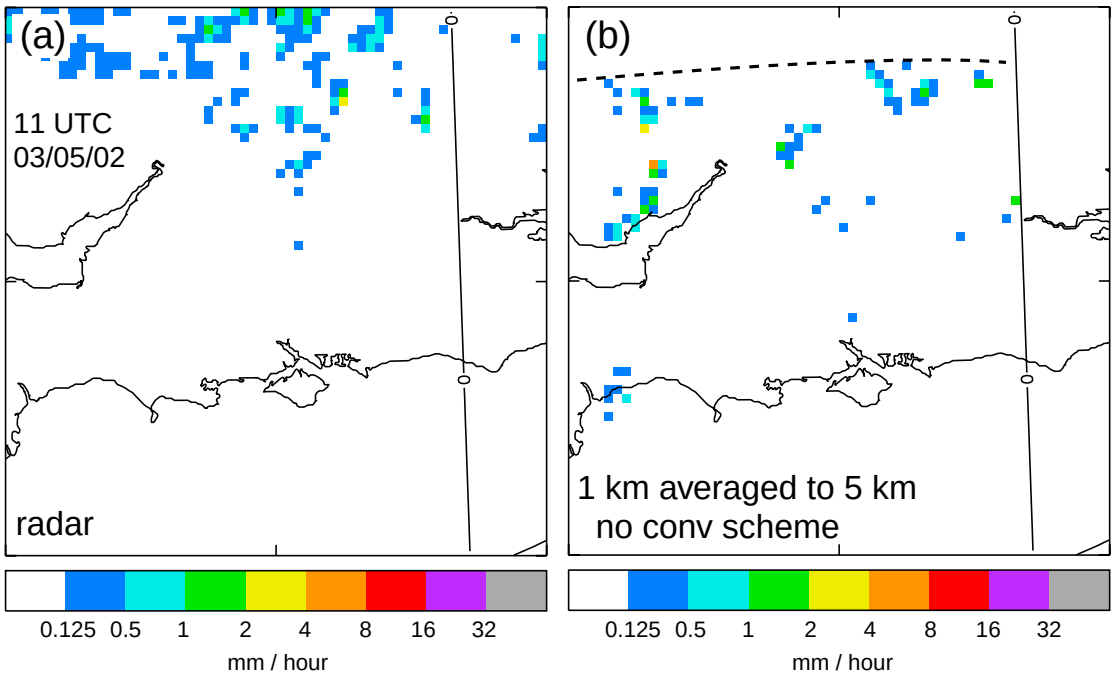


Figure 19. (a) Radar picture at 11 UTC 03/05/02 (Identical to Figure 18(a)). (b) Rain rates from a 1-km gridlength forecast averaged to the same 5km grid as the radar. The 1-km forecast started at 09 UTC following a 4-km forecast that started at 01 UTC.

The simulations that were closest to the radar at 11 UTC were the 4km with the convection scheme and the 1km with no convection scheme. They were the best forecasts for that time. When deciding whether one forecast was better than another, the exact locations of the showers were not under scrutiny because it is unreasonable to expect a 9-hour forecast to get that exactly right. It is the pattern of coverage and intensity that has been the focus of attention over the period surrounding 11 UTC.

5.2.2 14 UTC. Afternoon showers.

Rainfall rates at 14UTC from the five forecasts and the radar are shown in Figure 20. Showers had developed widely by this time and were heavy with lightning. The showers have the impression of being organised and not particularly heavy in the 5km radar picture, but the 2km image reveals that the cells were actually smaller, more intense (>32 mm/hour) and more scattered, though still with some organisation. The characteristics of each of the forecasts will now be described.

12 km with convection scheme – The spatial coverage of the showers was good and some organisation was evident. The intensities were lower than the radar but given that these are averages over 12 km squares and therefore ought to be lower it is a reasonable forecast. A 12km forecast will never reproduce the intensities of small showers, but statistical post-processing could help. All of the precipitation came from the convection scheme.

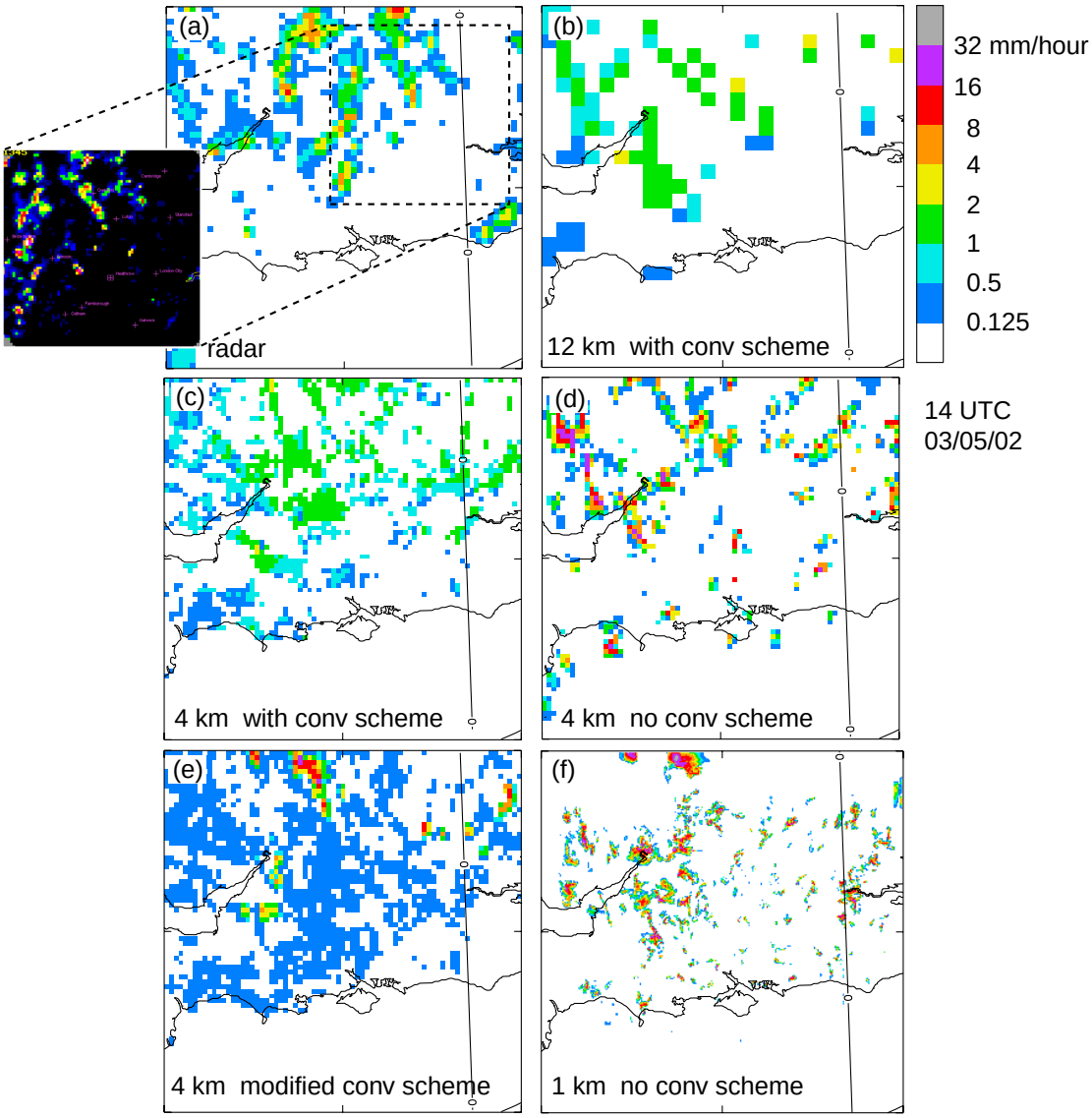


Figure 20. (a) Radar picture at 14 UTC 03/05/02. (b) to (f) Rainfall rates at 14 UTC from five different forecasts as indicated in the figure and in the text.

4 km with convection scheme – Showers were widespread with some organisation. The rain rates were clearly too low. They did not exceed 4 mm/hour when at that resolution they should have been > 8 mm/hour. The rates were even less than the 12-km forecast. All of the precipitation came from the convection scheme.

4 km with no convection scheme - Widespread showers had developed that were very intense (rates > 32mm/hour) in places. Some of the showers looked unphysical in the way they sometimes aligned themselves into single gridlength lines. When averaged to the same resolution as the radar (Figure 21(b)) the precipitation was still too heavy.

Unrealistically high rainfall rates (>100mm/hour) occurred in parts of the 4km domain not shown here.

4 km with modified convection scheme - Precipitation covered much of the area. Most of this was light with rain rates < 0.5 mm/hour and came from the convection scheme. Other regions with heavier precipitation were also present and they came from the model dynamics rather than the convection scheme. The rain rates in the heavier cells were too high even when averaged to 5km (Figure 21(c)) but did not reach the extreme values that occurred in 4 km run with no modification to the convection scheme.

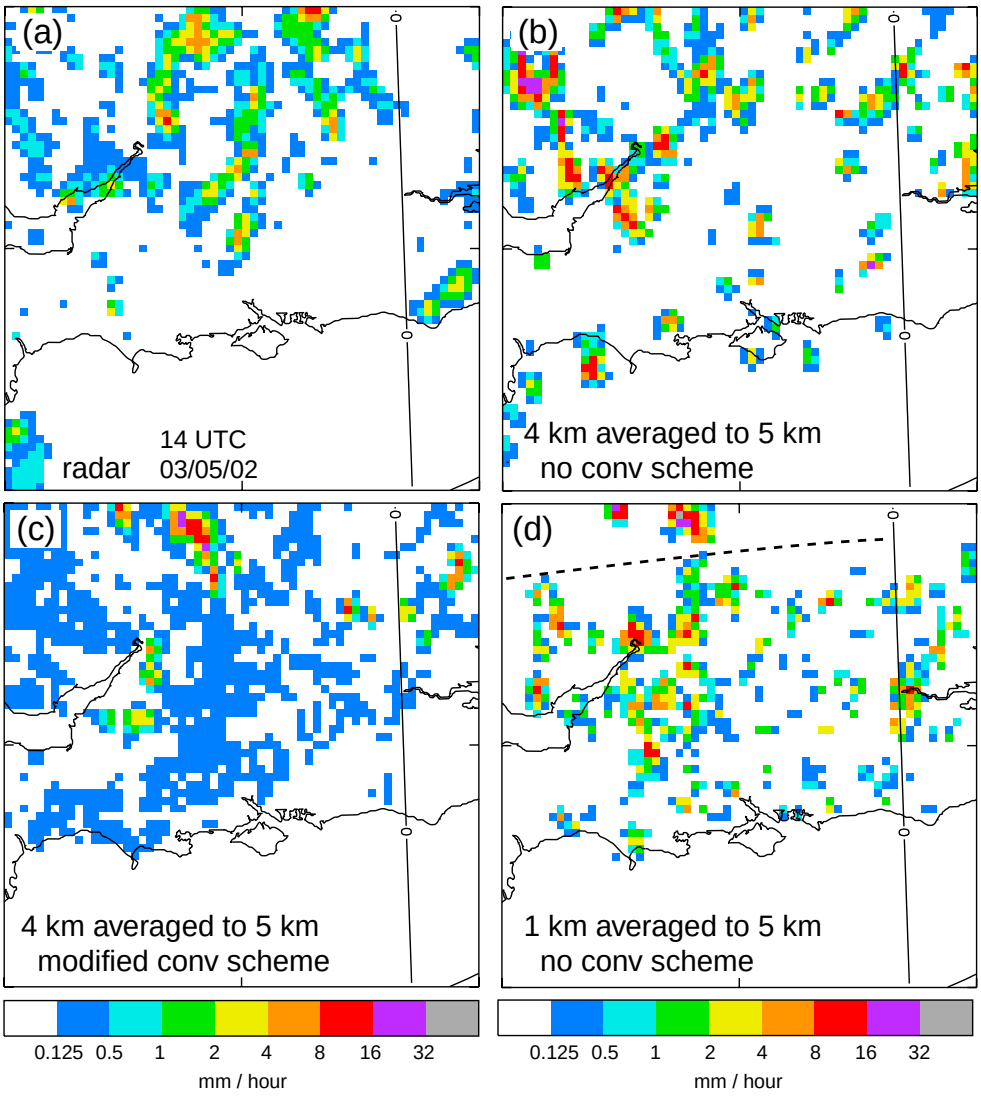


Figure 21. (a) Radar picture at 14 UTC 03/05/02 (Identical to Figure 20(a)). Rain rates averaged to the same 5km grid as the radar at 18 UTC from (b) 4 km gridlength forecast with no convection scheme, (c) 4 km with the convection scheme modified and (d) 1km gridlength scheme with no convection scheme.

1 km with no convection scheme - Showers had developed widely with rainfall rates exceeding 32 mm/hour. The showers appear to be too small and intense, but when averaged to 5km (Figure 21(d)) they look much closer to the radar at 5km, though still too scattered. The rainfall rates of >32 mm/hour were also seen on 2km resolution radar images and are therefore reasonable – though still too high. The empty strip at the north of the domain was still there, but less obvious because of the presence of large showers that

advected into the domain from the 4-km gridlength forecast that provided the boundary conditions.

All of the forecasts produced a good indication of convective activity in the region at 14 UTC. The 1 km gridlength forecast was the only one that was capable of generating showers with rainfall rates > 32 mm/hour and still be closest to the radar when averaged to 5km.

5.2.3 18 UTC. Evening organisation.

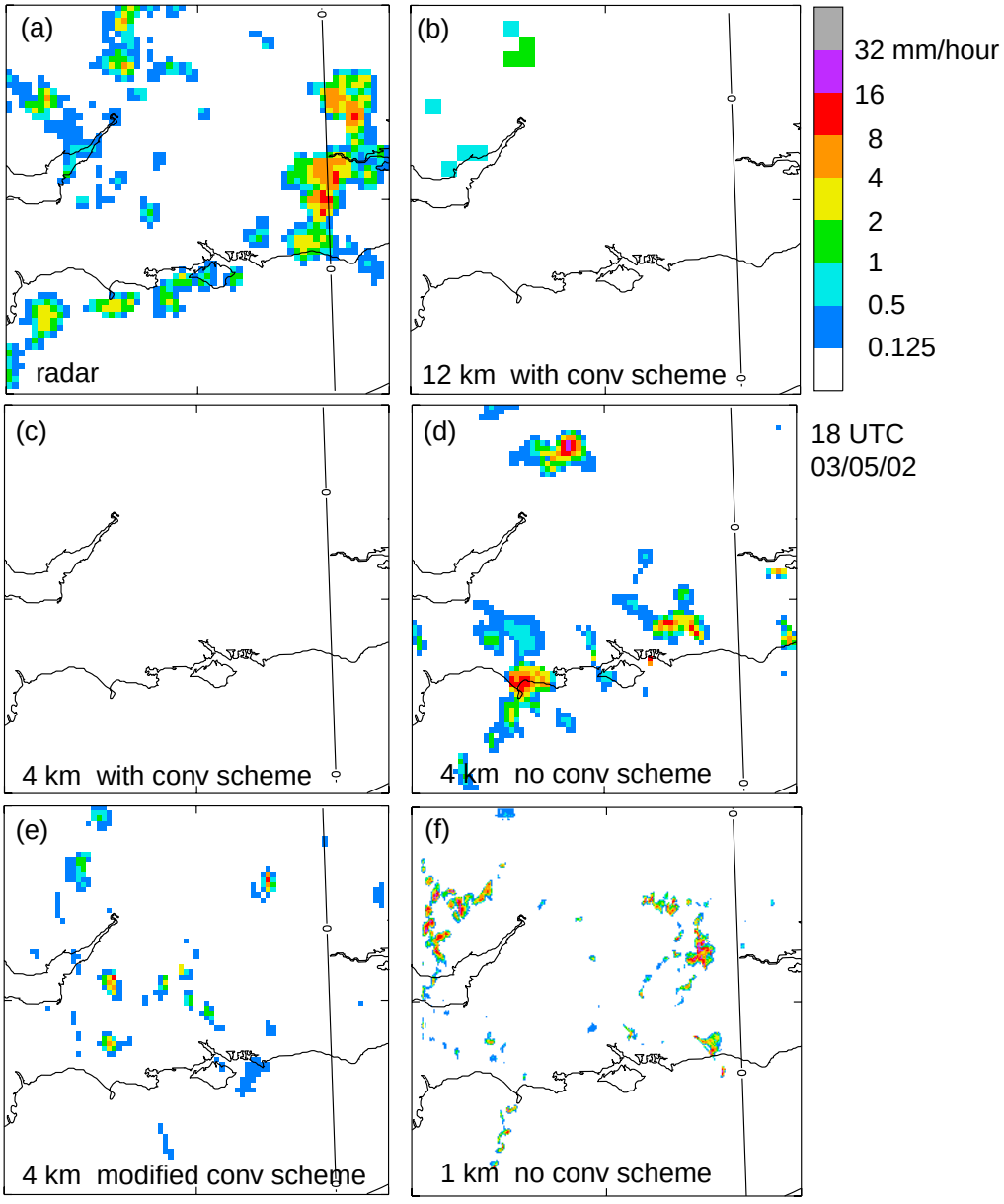


Figure 22. (a) Radar picture at 18 UTC 03/05/02. (b) to (f) Rainfall rates at 18 UTC from five different forecasts as indicated in the figure and in the text.

By 18 UTC the showers had organised into a band of thunderstorms that were moving through London and had extended into the Channel. Some other showers remained to the northwest of the area. It is at this time that the differences between the forecasts are most noticeable (Figure 22).

12 km with convection scheme – Nearly all of the convection had died out. No organised band developed. What precipitation there was came from the convection scheme.

4 km with convection scheme – All showers had died out.

4 km with no convection scheme – Convection had organised into larger showers. The size and intensity of the showers was very similar to the radar even though locations were not exactly correct (Figure 23(b)).

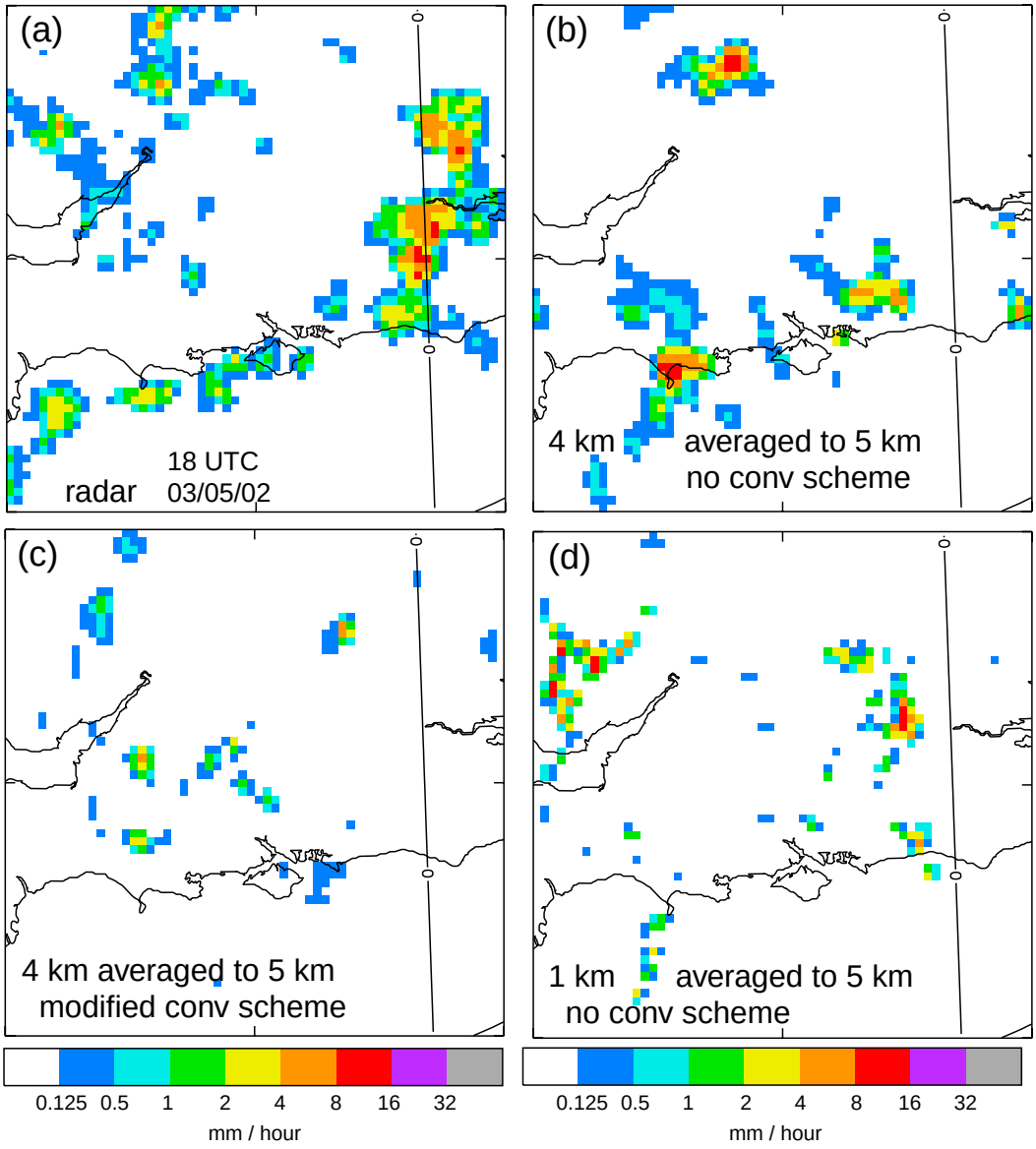


Figure 23. (a) Radar picture at 18 UTC 03/05/02 (Identical to Figure 22(a)). Rain rates averaged to the same 5km grid as the radar at 18 UTC from (b) 4 km gridlength forecast with no convection scheme, (c) 4 km with the convection scheme modified and (d) 1km gridlength scheme with no convection scheme.

4 km with modified convection scheme – Showers remained. They were not as heavy or as large as they should have been (Figure 23(c)) although some organisation had occurred. Shower location was poor. All of the precipitation came from the model dynamics rather than the convection scheme.

1 km with no convection scheme – The convection had organised into larger showers, but not to the extent that they should have. Figure 23(d) shows that the rainfall rates were close to the radar values but not widespread enough. Again, there is a rainfall void in the northern-most 30 km of the domain.

Only three of the forecasts had maintained any convection by 18 UTC. The 4 and 1-km forecasts with no convection scheme generated showers closest to the correct size and intensity. They are both a vast improvement on the 12-km forecast which could not maintain any showers once the solar heating became weaker because the convection scheme is not capable of organising convective cells.

5.2.4 Mean precipitation rates over the whole period

The differences between the five different forecasts over the whole period can be better highlighted by examining a graph of the mean rainfall rates (including zero points) within 1-km domain against time. This is plotted in Figure 24.
Significant shower activity developed at around 10:30 UTC. Both the 4 km run with the convection scheme and the 4km run with the modified convection scheme were around 15 minutes early, the 1 km and 12 km runs were around 30 minutes late and the 4 km with no convection scheme was an hour late.
In the period up to 13:00 UTC the convective activity increased. The 4-km forecast with the convection scheme and the 1-km forecast had the closest mean rates to the radar. After 13:00 UTC the convective activity levelled off until 15:00 UTC. Over this period both the 1 and 4 km forecasts with no convection scheme produced too much rainfall (especially the 4 km). The 4 km with the convection scheme remained the closest to the radar and the 4 km with the modified convection scheme improved.
After 15:00 UTC the shower activity increased again even though the solar heating was weakening. This was when the organisation of the showers became important. The 12-km run along with the 4km with the convection scheme could not reproduce the convective organisation and the mean rates dropped away. The 4-km forecast with the modified convection scheme was better because it maintained shower activity until 16:00 UTC before also dropping away. In contrast, the 1-km forecast maintained convective activity until around 16:30 UTC and the mean rates were almost the same as the radar between 15:30 and 16:30 UTC. Even by 18:00 UTC, the 1-km forecast still had significant activity, though half the radar value. The 4-km forecast became to active with almost twice the mean rates of the radar up to 16:00 before it too weakened dramatically for the rest of the forecast and ended up close to the 1 km forecast by 18:00.

This case has clearly shown the difficulty of producing accurate forecasts of the diurnal cycle of convective activity in situations when the showers are small and intense and organisation takes place. The only forecasts that could reproduce this organisation were the ones that did not use a convection scheme. In the 4-km run, this was at the expense of triggering the convection too late and producing a mean rain rate that was far to high. The 1 km forecast was much better and provides hope that it will be possible to simulate this sort of event much more realistically in the future. Figure 25 shows that the showers did indeed get larger with time in the 1-km run – there is a marked shift towards the right in the second plot. In future work, model output will be compared with high-resolution radar data in this way.

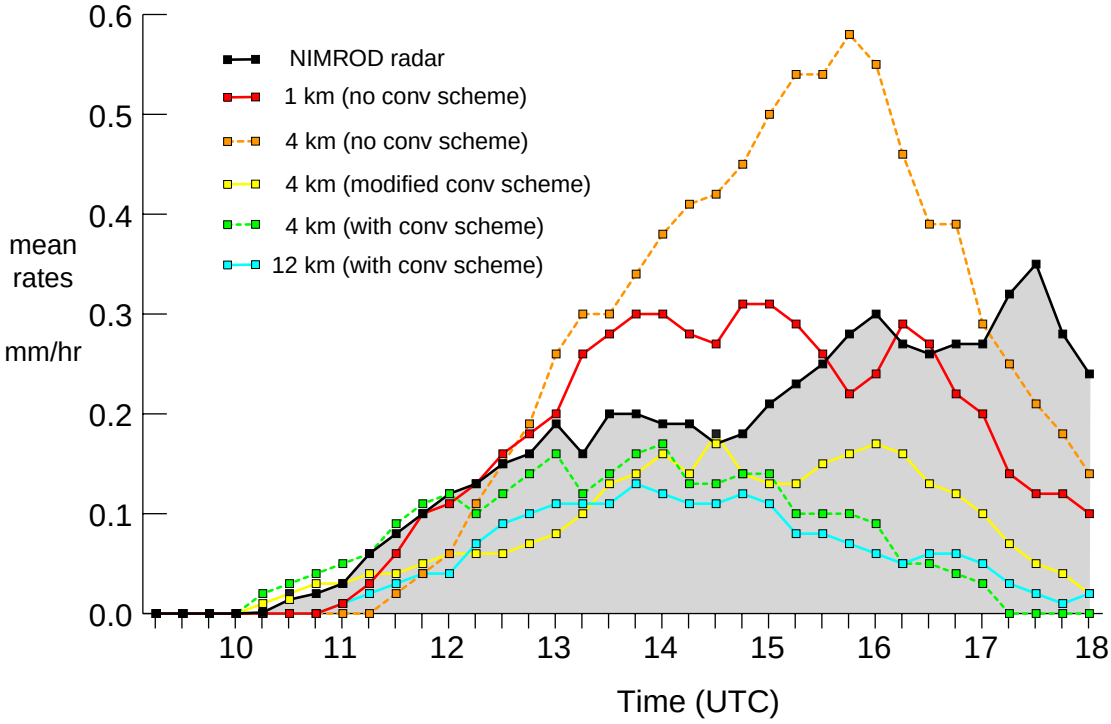


Figure 24. Graph of the mean rainfall rates within the 1 km domain against time for each of the forecasts described in the text (coloured lines) and the radar (black line enclosing grey shading).

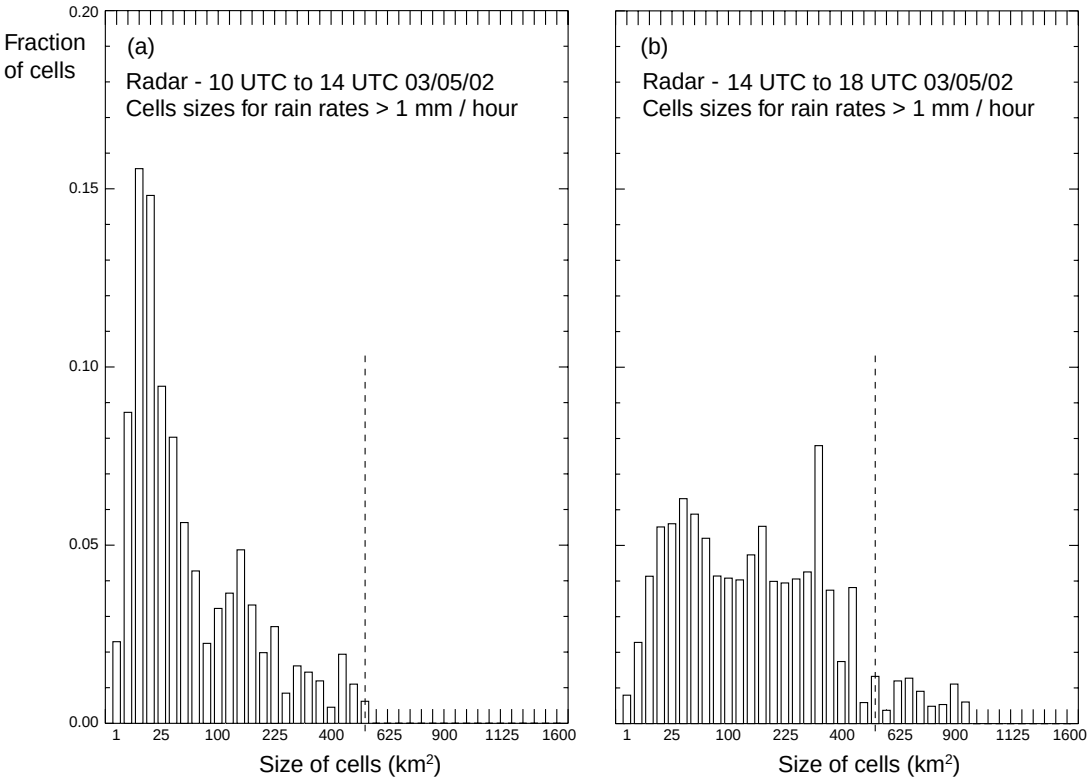


Figure 25. Histograms of the proportion of showers of different sizes, using 1 mm/hour as the threshold for a shower, from the 1-km gridlength forecast without the convection scheme (see text). (a) 10 to 14 UTC (b) 14 to 18 UTC 03/05/02.

5.2.5 Output of precipitation rates

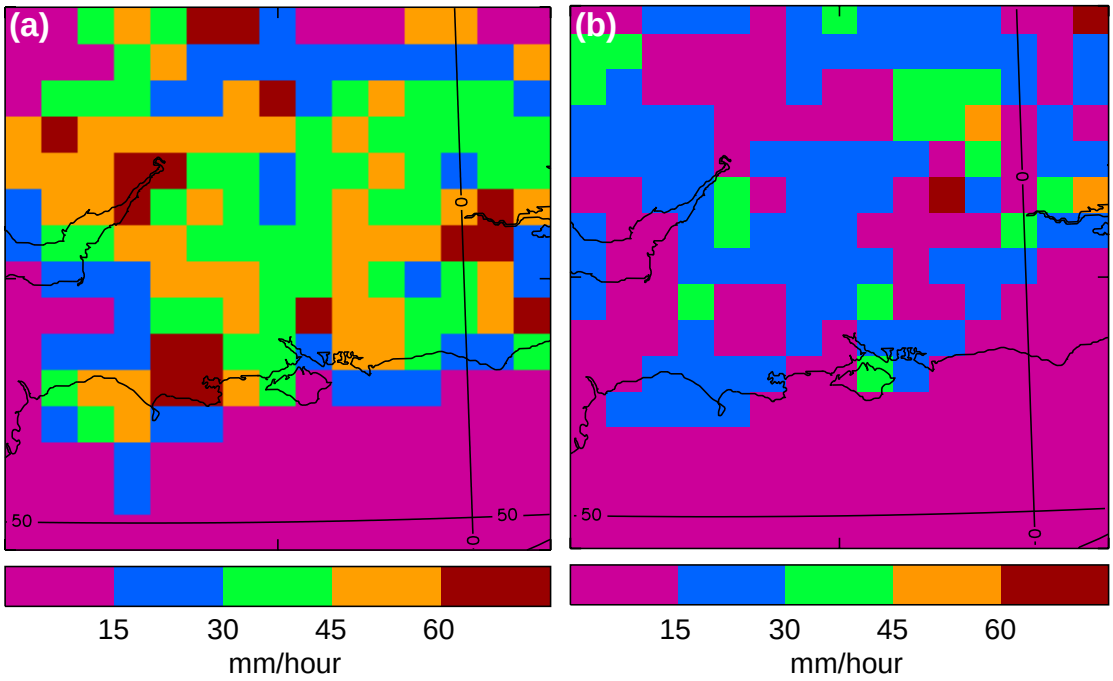


Figure 26. Peak rain rates within 20x20 km squares from 10 to 18 UTC 03/05/02 sampling at 5 minute intervals from (a) 1 km gridlength forecast with no convection scheme, (b) 4 km gridlength forecast with no convection scheme. (See text for more information about the forecasts).

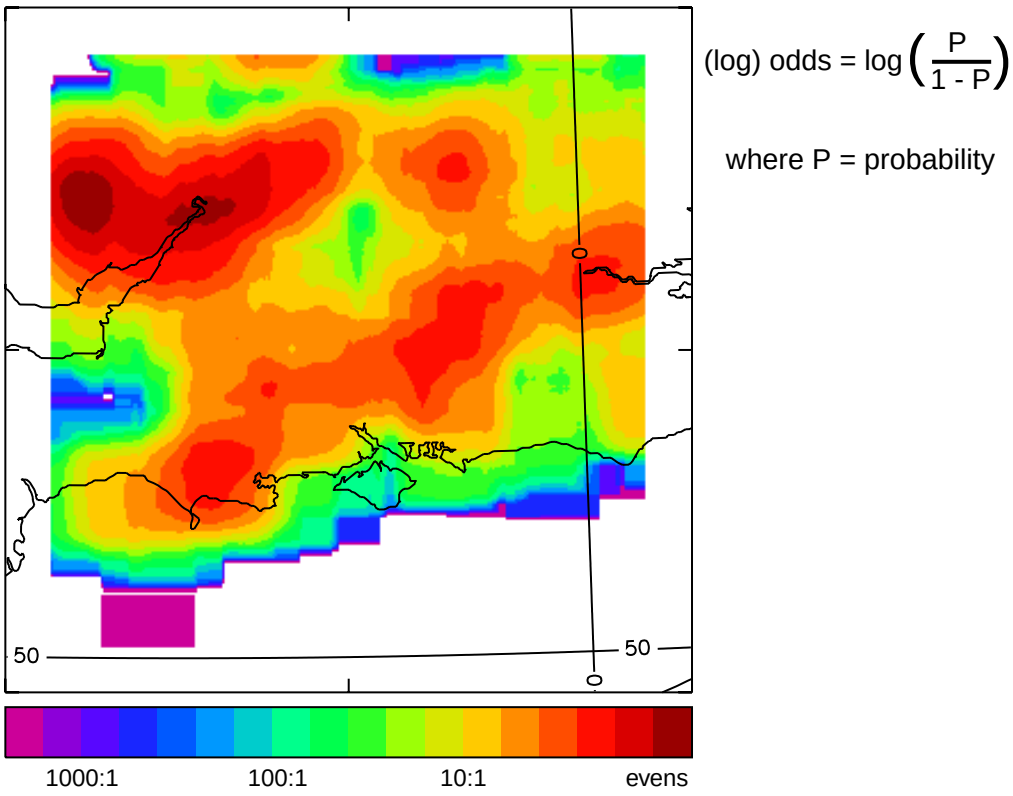


Figure 27. (log) odds of rain rates above 16 mm/hour from 'probabilities' generated from the fraction of gridpoints above 16 mm/hour within 40x40 km squares (see case study 2). Sampling was every 5 minutes within the period 10 to 18 UTC 03/05/02.

Even though the 4-km simulation without the convection scheme produced by far the highest mean precipitation rates and some extreme peak rates, it was not capable of generating the consistently high peak rain rates that the 1-km forecast produced. Figure 26 highlights this difference. The benefit of presenting the information in this way is to show how a 1-km forecast could be used to provide warning of the high rainfall rates that were observed. Another way of displaying rainfall rates from the model is shown in Figure 27. The probabilities were found using the method outlined in case study 2 before being converted into (log) odds. What this shows is the conditional odds for rain rates >16mm/hour given the forecast. (It is related but different to the (log) odds ratio used for verification in the previous case study which compared the conditional odds of a making a good forecast (a hit) to the odds of making a bad forecast (a miss)). If Figures 26 and 27 are used together, they reveal both the chances of getting significant rain rates along with the highest rainfall rates that could occur.

5.3 Summary of this case

The case was of scattered showers and thunderstorms that developed in the late morning and became organised into a larger band of storms during the evening. The model was run with gridlengths of 12, 4 and 1 km. Five simulations were presented. The 1-km gridlength forecast with no convection scheme was the best overall. It produced the most realistic magnitudes of peak and average rain rates as well organising and maintaining convection into the evening.

This case has also highlighted some issues that need addressing. The inhibition of shower formation in the 1-km forecast, close to the northern boundary where flow was coming into the domain, is of concern. Further work is required to examine ways of alleviating this problem, and in particular, the impact of using different 4km gridlength forecasts to supply the boundary conditions. The use of a modification to the convection scheme to limit its effect in a 4km forecast has had a small positive impact, but on the basis of these results, further tests to limit the convection scheme even more at 4 km should be tried. The length of the forecasts may have been important; shorter-range forecasts with better initial conditions should be run to examine whether the same conclusions can be drawn. An investigation of what happens after 18 UTC should also be carried out.

Criteria for forecast assessment	The best forecast
Magnitude of precipitation rates	1 km, no convection scheme
Initiation time	4-km, with convection scheme & 4-km with modified convection scheme.
Mean precipitation rates	1-km, no convection scheme
Organisation of showers	1-km, no convection scheme & 4-km, no convection scheme

6 Case study 4. 29th July 2002

6.1 Description of the event

An isolated cluster of thunderstorms developed over Essex during the early hours of the morning and tracked northwards (Figure 28). Rainfall rates were very high with localised flooding reported. At the time, England and Wales were under the influence of a gentle flow of hot air from the continent.

The triggering was not associated with daytime heating because the event occurred overnight. Neither was it likely to have been the result of sea-breeze convergence or an urban heat island effect. It would also appear that orography played no part, as the area was flat. An explanation of why the storms developed in that general area can be provided by looking at satellite water vapour imagery to see what was happening in the middle and upper troposphere.

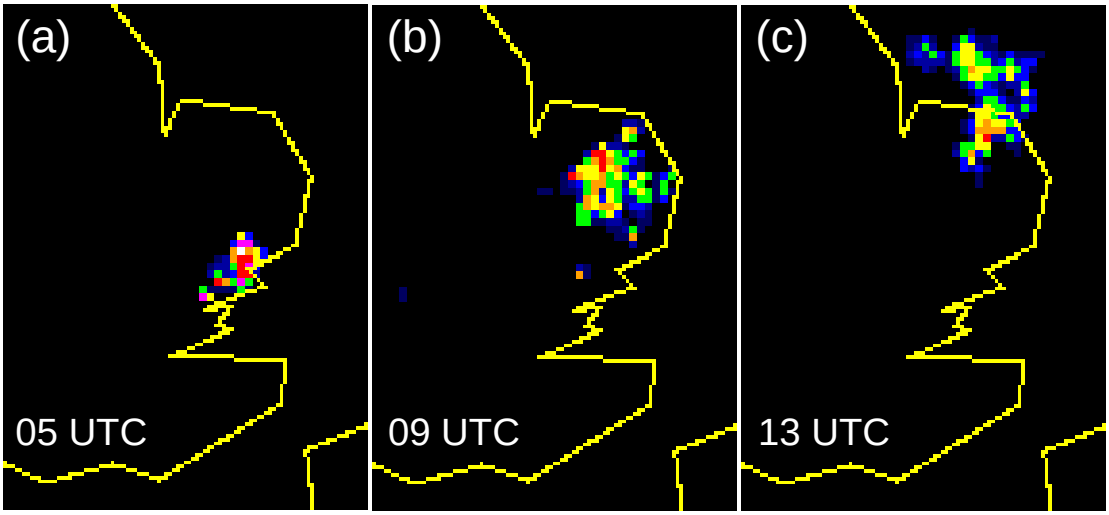


Figure 28. Rainfall rates in mm/hour from radar for 05, 09 and 13 UTC 29/07/02. Colours are: dark blue > 0.125, lighter blue > 0.5, green > 1.0, yellow > 2.0, orange >4.0, red > 8.0, purple > 16.0 & white > 32.0 mm/hour.

The water vapour image shown in Figure 29 reveals that there were vortices present in the upper troposphere. Such features are sometimes associated with regions of thunderstorm development as they provide the instability and ascent that is required (Roberts 2000). In this case, the storm did appear to develop in the expected region along the edge of the dry (blue) zone where mid-tropospheric ascent would probably have been occurring. An accurate forecast of this storm would require both the upper-tropospheric vorticity anomaly and the local triggering mechanisms to be modelled correctly. An examination of global model (~60km gridlength) output (not shown) revealed that even at such coarse resolution there was a vorticity anomaly present albeit somewhat misplaced and too weak. It would seem then, that the model (at all resolutions) had favourable conditions for convective development in the region of interest. The aim was to see whether changing the resolution would have a significant impact on convective triggering and evolution.

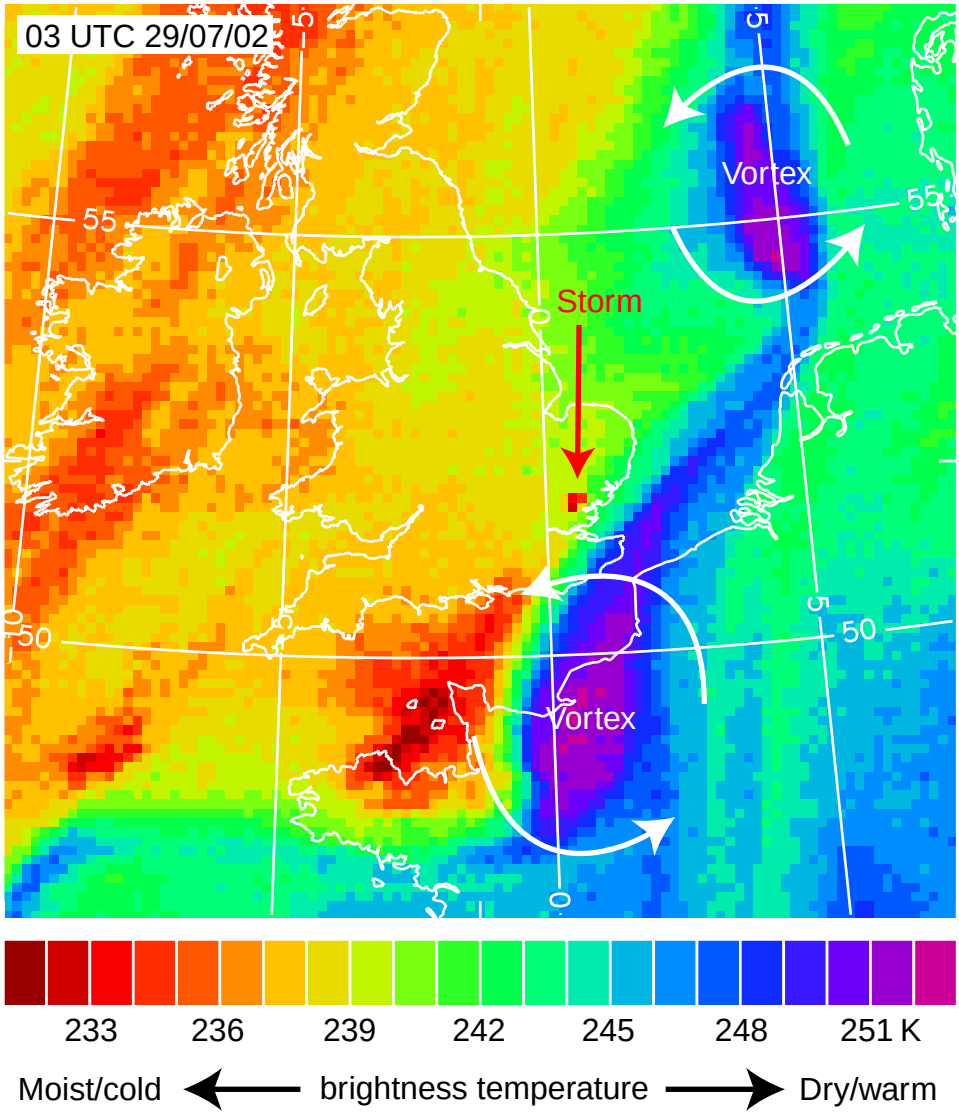


Figure 29. Meteosat water vapour image for 03 UTC 29/07/02. Two upper-tropospheric mesoscale vortices are highlighted with white arrows. The red arrow points to the early development of the thunderstorms.

6.2 Comparison of forecasts using different resolutions

Model simulations with the new-dynamics were run with gridlengths of 12, 4 and 1.2 km and 38 or 76 vertical levels. All the forecasts started from 01 UTC 29/07/02 and ran for 12 hours ending at 13 UTC. The 4 and 1.2 gridlength simulations used the fields interpolated from 12km at 01 UTC. No extra observational information was added to any of the simulations. The domains used are shown in Figure 30.

A large number of model simulations were run and examined. Some are listed below. Not all will be discussed.

1. 12 km gridlength, 38 levels, convection scheme switched on. New operational formulation.
2. 4 km gridlength, 76 levels, convection scheme switched on
3. 4 km gridlength, 76 levels, convection scheme switched off

4. 4 km gridlength, 38 levels, convection scheme switched on
5. 4 km gridlength, 38 levels, convection scheme switched off
6. 4 km gridlength, 76 levels, modified convection scheme
7. 4 km gridlength, 38 levels, modified convection scheme
8. 4 km gridlength, 76 levels, modified convection scheme, change to the critical relative humidity values
9. 1.2 km gridlength, 76 levels, convection scheme switched off
10. 1.2 km gridlength, 38 levels, convection scheme switched off

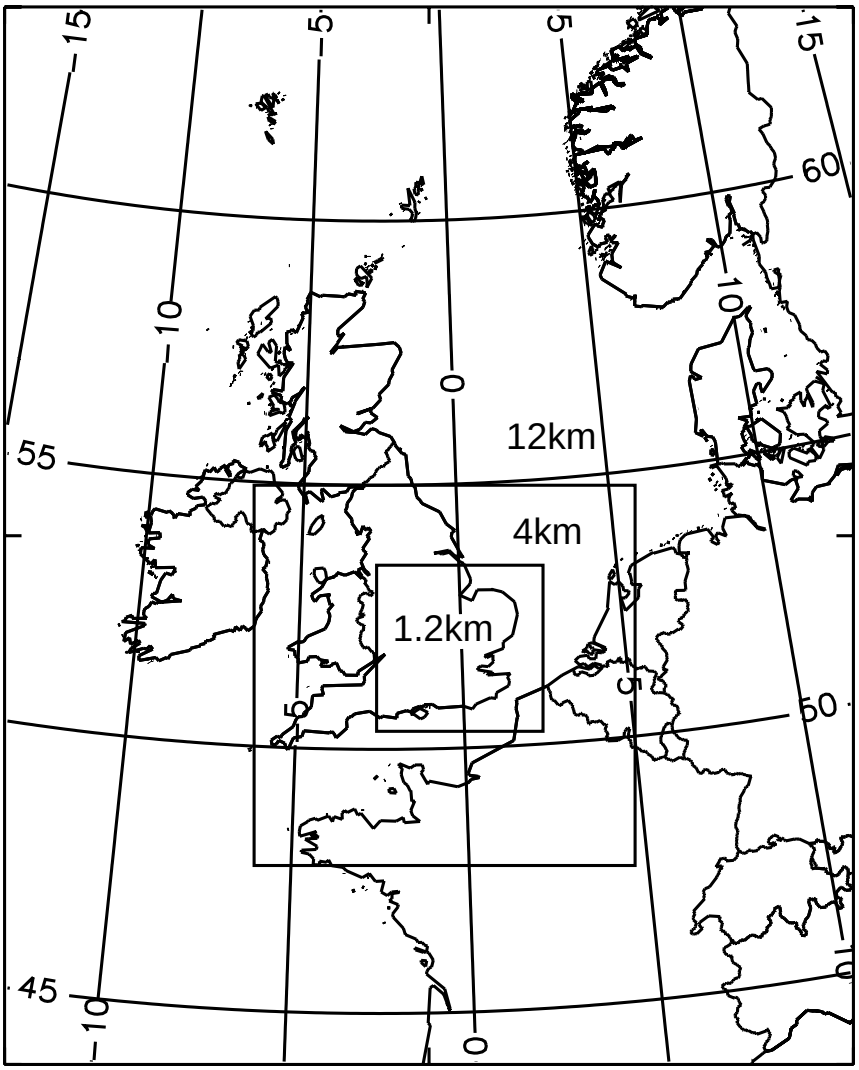


Figure 30. The domains used for the 12, 4 and 1.2-km gridlength simulations

6.2.1 The 12 and 4 -km gridlength simulations

The 12km gridlength forecast from 00 UTC and one of the 4km forecasts are shown in Figure 31. The 12-km, 38-level forecast did produce light precipitation for a while, but it died out and no storm system developed. All of the rain came from the convection scheme. The 4-km, 76 level simulation generated precipitation within the same period and over almost exactly the same area as the 12km forecast. It also had the same light intensity and was generated entirely from the convection scheme. However, the precipitation did not completely vanish in the 4-km forecast, a few isolated gridpoints had rain from the convection scheme but no organised system formed.

The 4-km run with 38 levels was very similar to the 76 level run. The 4-km runs with no convection scheme (38 and 76 levels) developed no precipitation at all.

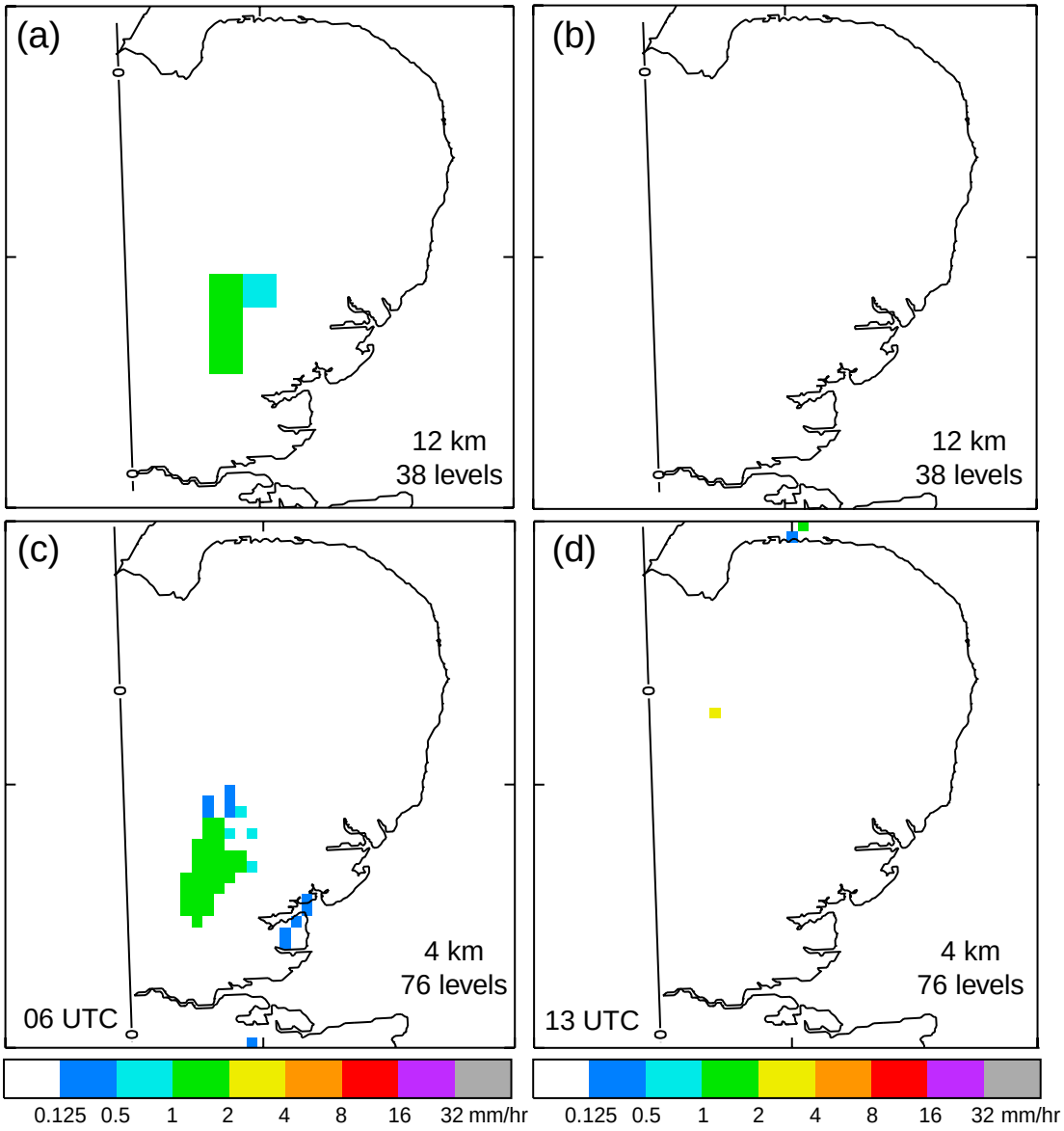


Figure 31. Rainfall rates at 06 and 13 UTC 29/07/02 from forecasts starting at 01 UTC. (a) & (b) 12-km gridlength, 38 level forecast. (c) & (d) 4-km gridlength, 76 level forecast. Both forecasts ran with the convection scheme switched on.

6.2.2 The 1.2-km gridlength simulations

The model was run with a gridlength of 1.2 km using either 38 or 76 vertical levels. Both of these simulations did generate some high rainfall rates (Figure 32) but were significantly different from each other. The 76 level forecast developed some small scattered cells with high precipitation rates >32 mm/hour later in the forecast (after 07:00), whereas the 38-level run triggered earlier and developed a larger cluster of precipitating grid-squares that also had rates > 32 mm/hour. Not only did this 38-level, 1.2-km gridlength simulation develop a region of organised heavy showers but it generated them to within 30 km of

where they actually occurred – and 12 hours into the forecast. This is a much better forecast than any of the other simulations and an encouraging result because it shows that going to a 1km gridlength can have a positive impact on the triggering of convection in what appears to be a marginal situation.

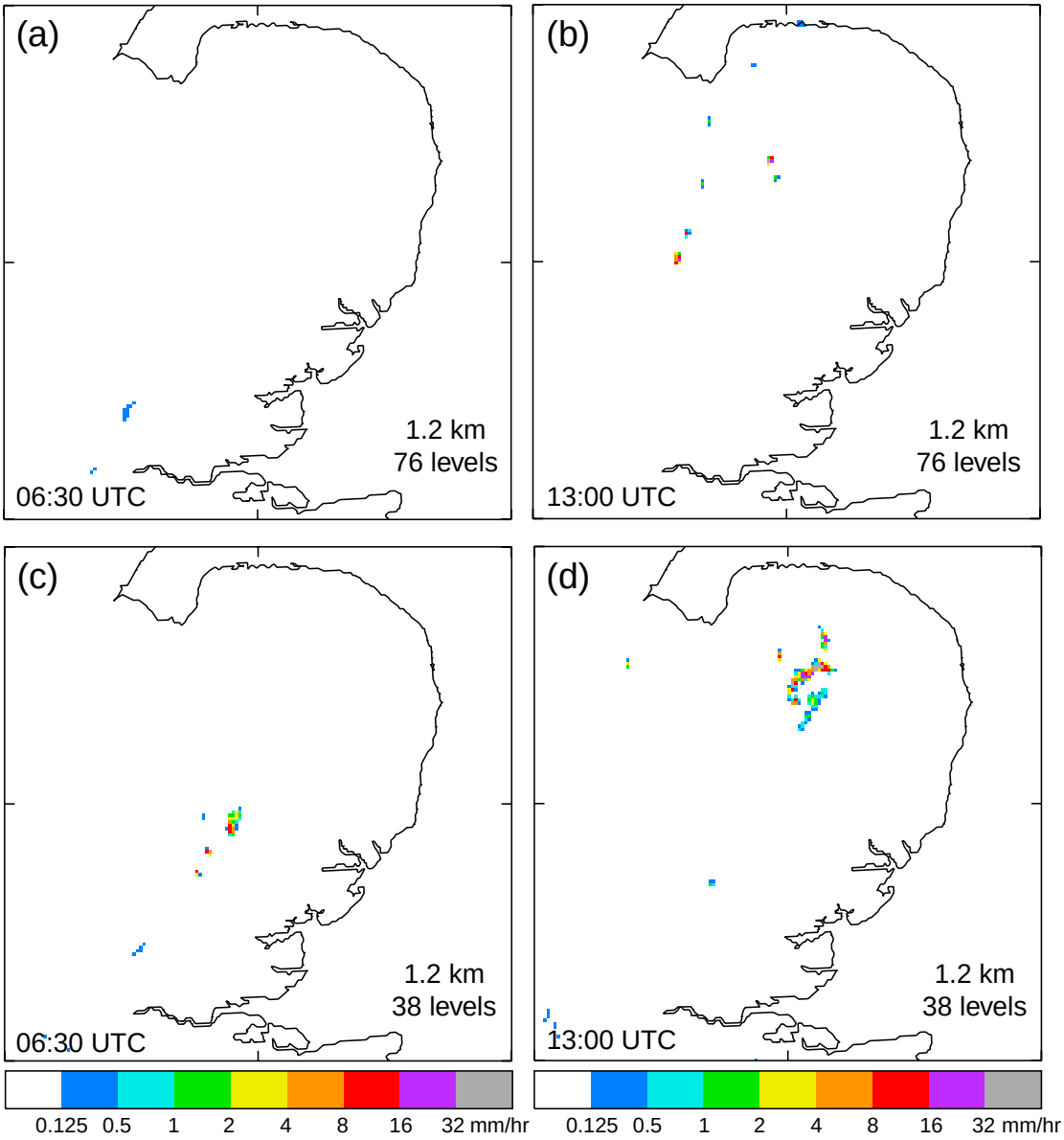


Figure 32. Rainfall rates at 06:30 and 13:00 UTC 29/07/02 from 1.2 km gridlength forecasts starting at 01:00 UTC using (a) & (b) 76 levels, (c) and (d) 38 levels.

Despite the success of the 1-km, 38-level simulation, it is curious that the 38 and 76 level forecasts were so different. This will be investigated, but is not yet understood. It is possible to speculate that the extra resolution in the 76-level run had enabled a stronger capping inversion to form and thus provided extra inhibition to any convective triggering. Careful analysis of the sensitivity of triggering to vertical resolution is required. The current consensus is that it would be better to run high-resolution simulations with 76 levels rather than 38 so that there is a more sensible ratio between the horizontal and vertical grid spacings (Lean and Clark 2003) – but convection experiments may yet contradict this view.

Figure 33 shows how close the 38-level forecast was to the radar over the period 12 to 13 UTC in terms of both the magnitudes of the accumulations and the location. The distance between the centre of the observed and forecast storms at this time was only 20 km. This is impressive for a 12-hour forecast, but does over-emphasise the success somewhat. When accumulations over a longer period are examined, the results are not so good. The model simulation did not produce any earlier hourly accumulations greater than 2 mm, whereas the hourly radar accumulations exceeded 12 mm. So, even though the 38-level, 1-km run did eventually generate heavy rain in nearly the correct location, it did not develop enough precipitation or trigger early enough. It may be that part of the reason for this is due to errors in the larger-scale forcing in the model (misplaced upper-level vortex) and this requires investigation. Experiments that examine the sensitivity of the forecast to changes in the initial conditions might provide further insight and be the key to producing better simulations.

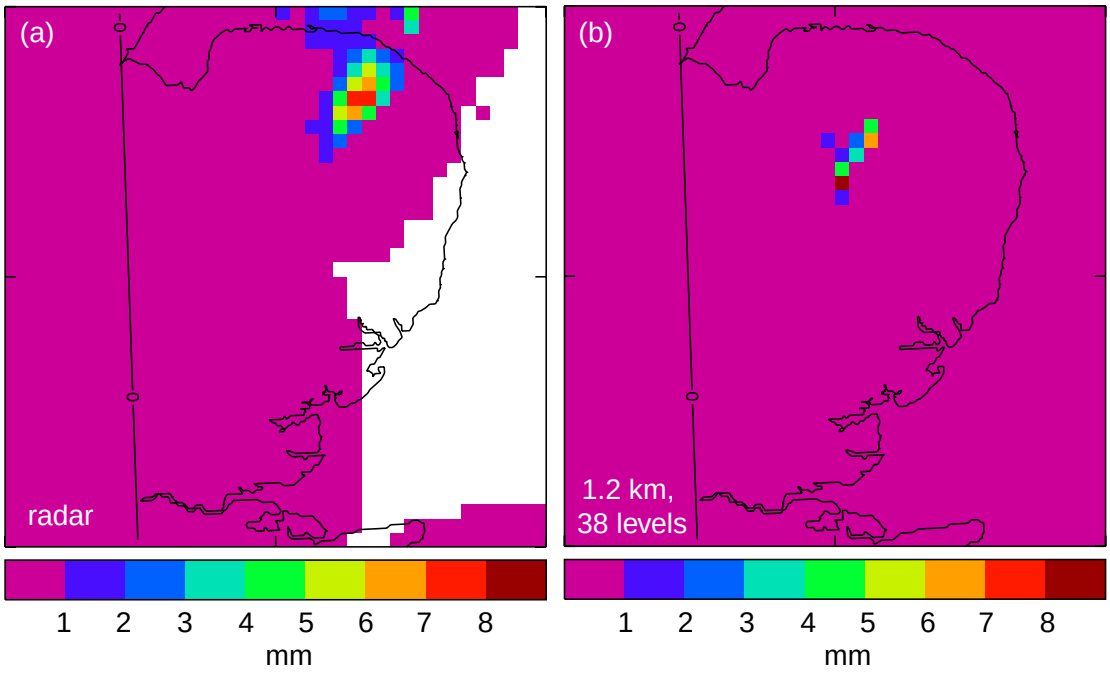


Figure 33. 1-hour rainfall accumulations from 12 to 13 UTC 29/07/02 from (a) NIMROD processed radar and (b) 1.2 km, 38-level forecast starting from 01 UTC.

6.2.3 The 4-km gridlength, 76 level run with modified convection scheme

The 4-km forecast that used a modification to limit the effect of the convection scheme is shown in Figure 34. The interesting thing about the precipitation in this run is the way it was distributed in an arc shape stretching from northern France through to southeast England (Figure 34(a)). All of this rain came from the convection scheme and is light because of the restriction on the scheme. It marked a region where there was just enough instability for the convection scheme to trigger and was arc shaped because it was related to the pattern of larger-scale instability in the model. Although the convection scheme triggered, the resolution was too coarse for the model dynamics to develop any showers. The difference between this run and the 1-km, 38-level forecast is that the 1-km run did manage to trigger its showers within this arc. The cluster of showers in Figure 32(c) are located in the same place as the area marked 'A' in Figure 34(a). This is also the same region where the 12 and 4km runs shown in Figure 31 produced rain from the convection scheme.

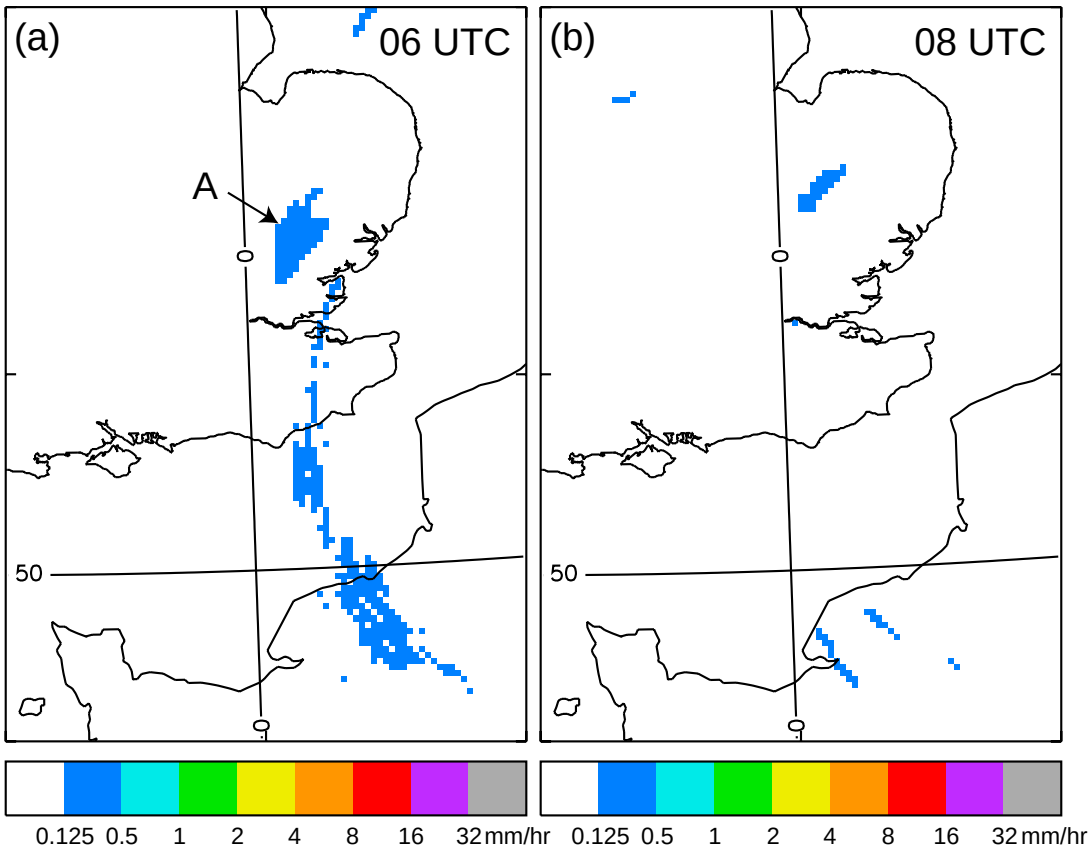


Figure 34. Rain rates at 06 and 08 UTC 29/07/02 from a 4-km gridlength, 76 level simulation starting at 01 UTC. The label 'A' is discussed in the text.

6.2.4 Output of precipitation rates

Figure 35 reveals very clearly the differences between the two 1.2 km gridlength simulations. It is clear that the 38-level run was much better because it produced high rainfall rates within the region where the storms occurred. The 38-level simulation generated rainfall rates that exceeded 100 mm/hour. These values may seem high, but given that 5km averaged radar pixels exceeded 32 mm/hour and that flash floods did occur and that these rainfall rates are typical of such storms, it is reasonable. If an operational situation were imagined, output of this kind would have provided a clear warning signal that intense thunderstorms might occur in the region where they did. The current operational 12-km gridlength setup could give no such warning of this event.

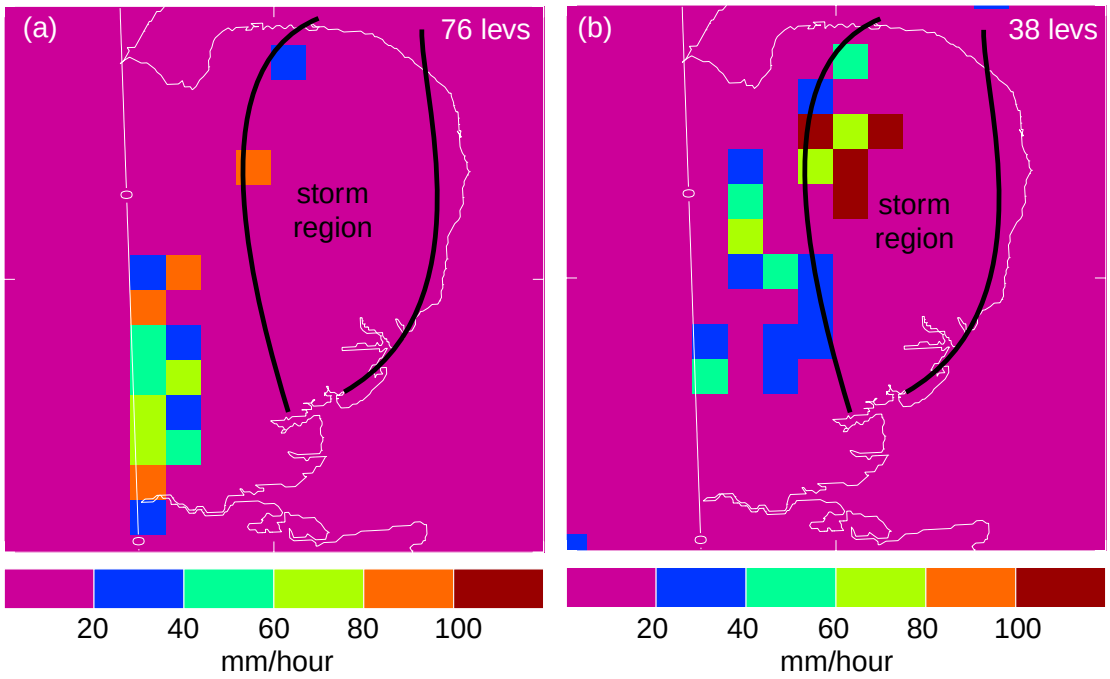


Figure 35. Peak rainfall rates within 20x20 km squares over the period 03 to 13 UTC 29/07/02, sampling at 5 minute intervals from 1.2 km gridlength simulations starting at 01 UTC. The runs used (a) 76 and (b) 38 vertical levels.

6.3 Summary of this case

The case was of an isolated cluster of thunderstorms that produced heavy rain and localised flooding. The model was run with gridlengths of 12, 4 and 1.2 km and 38 or 76 levels. Of all the simulations run, only the 1.2-km runs triggered any heavy showers from the model dynamics rather than just precipitation from the convection scheme. It is not necessarily better to have produced resolved rather than parametrized convection, but in this case, a storm of the size of the one observed should be resolved (and therefore triggered) at both 1.2 and 4km. The 1.2-km, 38 level simulation produced storms of realistic intensity and within 30 km of where they occurred. Although by no means a 'perfect' forecast, it was a significant improvement on the coarser resolution forecasts and highlights the potential of storm-scale numerical modelling.

Criteria for forecast assessment	The best forecast
Convective initiation time	4 km, 38 or 76 levels with conv scheme 1.2 km, 38 levels
Magnitude of rainfall rates	1.2 km, 38 levels
Magnitude of rainfall accumulations	1.2 km, 38 levels
Warning of severe storms	1.2 km, 38 levels
Spatial accuracy	1.2 km, 38 levels

This case is ideal for further tests of the sensitivity of model simulations to changes in resolution, parameters, formulation and initial conditions as well as for examining some of the mechanisms involved in convective triggering.

7 Conclusions and further work

7.1 Conclusions

The highest-resolution simulations produced the best results in all of the case studies.

Case study 1. The 2 and 4-km gridlength forecasts were capable of generating a squall line that propagated at the correct speed. The 2-km forecast produced the very high rainfall rates that were observed.

Case study 2. The 2-km gridlength simulation significantly improved both the position and magnitude of a band of high rainfall accumulations compared to the 12 and 4-km forecasts. The displacement error was reduced by half. Verification scores showed this improvement.

Case study 3. Only the 1-km gridlength simulation could maintain convection into the evening by organising scattered convection into larger clusters as well as producing a reasonable representation of the early afternoon showers.

Case study 4. The 1.2-km simulations managed to trigger convective storms when no other forecasts could. The 12-hour forecast with 38 vertical levels generated an isolated and intense storm only 20 km away from where the observed thunderstorm occurred.

These results are encouraging, more so, because no extra observations were added to the high-resolution simulations. They have provided examples of how a storm-scale modelling system has the potential to deliver more accurate short-range forecasts of high-impact precipitation events. Particularly pleasing is the way that the extra resolution can lead to an improvement in the triggering or location of precipitation and not just in the intensity. This is crucial for forecasting on the scales of small/medium sized river catchments and should improve further as techniques are introduced to include observations in high-resolution models.

It must be emphasised here, that this work is looking to the future. It is not possible to run forecasts with gridlengths of ~1km operationally at present. The computing resources required are too great. The time required to run a 12-km gridlength forecast for 12 hours was around 20 minutes; the 4-km simulations used 3 times as many computer processors and took around 1.5 hours; the ~1-km simulations used 4.5 times as many processors (as 12km) and took around 5 hours. To run a ~1km forecast in real time is not a practical consideration now, but should be a possibility in the future as computer power continually improves.

7.2 Future work

Whilst the results so far are encouraging there is still a considerable amount of research and development work required to end up with an operational storm-scale model forecast system that deals with the extra difficulties that high resolution brings. Some of the future work that is envisaged for this project as part of that effort will be outlined below. Several more case studies are planned. The work can be split up into three distinct areas: (1) Investigation of the impact of changes to model configuration, parameters and surface fields. (2) Investigation of the impact of changes to the initial state of high-resolution forecasts through the inclusion of observations or from random perturbations. (3) Development of appropriate diagnostic products for high-resolution precipitation forecasts along with methods of forecast verification.

1. Further experiments are still required to examine further the impact of modifications to the convection scheme in 4-km runs (and perhaps at 1 km). This should also have an effect on the 'spin up' of convection at the inflow boundaries of the 1-km domain which was shown to be of particular concern in case study 3. The sensitivity to other parameters in the model such as diffusion will also be examined along with tests new microphysics code that has been developed at the JCMM. It is expected that the accuracy of high-resolution forecasts will also be sensitive to domain size and this needs to be tested. Surface fields such as soil moisture are likely to play an important role in convective situations. A new soil moisture analysis based on Nimrod radar data has been developed at JCHMR and will be included.
2. In the case studies so far, observations are not assimilated into the model when gridlengths are finer than 12km. Experiments will be performed to initially examine the effect of changes to the initial fields at 12km gridlength. Then the intention is to look at the impact of assimilating observations at higher resolution using techniques being developed at JCMM. It is hoped that it will also be possible to examine the sensitivity of adding random perturbations to dynamic or moisture fields and build up an ensemble of simulations.
3. One of the major uses for a storm-scale modelling system will be the production of forecasts of precipitation over small or medium sized river catchments. To do this properly, precipitation diagnostics need to be designed that extract the extra information from high-resolution forecasts (such as peak rainfall intensities) but present only the spatial detail that is appropriate for the accuracy of the model and the size of the catchment. Such diagnostics need to be useful as input into hydrological models. In addition to designing new diagnostics, suitable methods of evaluating the performance of storm-scale models are required. Ideally these methods should be related to the new diagnostics, so that model evaluation is directly linked to products that are generated.

Up until now, the experiments have not used a single model configuration, as would be the case in an operational system. Each of the case studies has been run with a variety resolutions, domains and parameter settings. This was appropriate when the main aim was to see whether a high-resolution model (that was still under development) would run and produce reasonable results. A concern might be that some of the benefits shown in this report would not be so apparent if a single configuration were used for all the studies. This issue will be addressed in future by adopting a 'pseudo-operational' model configuration as a starting point for sensitivity experiments, which will be the same as the setup used in the high Resolution Trial Model (HRTM) project at the Joint Centre for Mesoscale Meteorology (JCMM) in Reading.

8 References

Cullen, M.J.P., Davies, T., Mawson, M.H., James, J.A., Coulter, S.C. and Malcolm A., 1997, An overview of Numerical Methods for the Next Generation UK NWP and Climate Model" Numerical Methods in Atmospheric and Ocean Modelling. The Andre J.Robert memorial volume. Edited by Charles A Lin, Rene Laprise and Harold Ritchie 425-444

Lean, H. W., 2001: A Study of the Effects of Horizontal and Vertical Resolution in the new Dynamics representation of Faxtex IOP16. *JCMM internal report.,* **132**. *(NWP Scientific Report.,* **64***). Q.J.R.Meteorol.Soc accepted?*

Lorenc,A.C., Ballard, S.P., Bell,R.S., Ingleby,N.B., Andrews,P.L.F., Barker,D.M., Bray,J.R., Clayton,A.M., Dalby,T., Li,D., Payne,T.J. and Saunders., F.W. 2000: The. Met Office Global 3-Dimensional Variational Data Assimilation Scheme, *Q.J.R.Meteorol.Soc.*, **126**, 2991-3012

Roberts, N. M., 2000: The relationship between water vapour imagery and thunderstorms. *JCMM internal report.,* **110**. *(NWP Technical Report.,* **300***)*

Stephenson, D.B., 2000: Use of the "odds ratio" for diagnosing forecast skill. *Wea. Forecasting,* **15***,* 221-232.

9 Acknowledgements

This work was partially funded by Defra under the Flood Forecasting and Warning Theme. I would like to thank Brian Golding for very useful comments on an earlier draft of this report. I would also like to thank Humphrey Lean for technical help and for valuable discussions about this work. Andrew Macallan provided essential computing support and generated satellite data in a useable format.